Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Context Effects in Judgment: Adaptation Level as a Function of the Mean, Midpoint, and Median of the Stimuli: A. PARDUCCI, R. C. CALFEE, L. M. MARSHALL, AND L. P. DAVIDSON 65

Stimulus Differences in Discrimination Learning: R. A. CHAMPION AND C. D. STANDISH 78

Additivity of Cues in Discrimination Learning of Letter Patterns: T. R. TRABASSO 83

Effect of Intermittent Reinforcement on Acquisition, Extinction, and Spontaneous Recovery of the Conditioned Eyelid Response: T. F. HARTMAN AND D. A. GRANT 89

Response Strengthening by Information and Effect in Human Learning: W. A. HILLIX AND M. H. MARX 97

Frequency and Usefulness of Verbal and Nonverbal Methods in the Learning and Transfer of a Paired-Associate Serial Motor Task: E. NEUMANN 103

Likelihood Judgments and Sequential Effects in a Two-Choice Probability Learning Situation: N. H. ANDERSON AND R. E. WHALEN 111

Woodworth Scale Values of the Lightfoot Pictures of Facial Expression: N. LEVY AND H. SCHLOSBERG 121

Tactual Form Discrimination: A Development Comparison under Conditions of Spatial Interference: E. S. GOLLIN 126

Effect of Self-Satiation on Perceived Size of a Visual Figure: C. P. DUNCAN 130

American Psychological Association

Vol. 60, No. 2 August 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 60, No. 2 AUGUST 1960

CONTEXT EFFECTS IN JUDGMENT: ADAPTATION LEVEL AS A FUNCTION OF THE MEAN, MIDPOINT, AND MEDIAN OF THE STIMULI [1]

ALLEN PARDUCCI, ROBERT C. CALFEE, LOUISE M. MARSHALL
University of California, Los Angeles
AND LINDA P. DAVIDSON
Swarthmore College

Perceptual judgments, particularly for simple sensory dimensions, are largely determined by the physical relationships among the stimuli presented for judgment. For comparative judgments, the most important relationship is between the stimuli singled out for comparison—usually the standard and whatever comparison or variable stimulus is present at the time of judgment. But other context stimuli such as the background and the preceding variable stimuli may also affect the judgment (e.g., Garner, 1954; Künnapas, 1955; Turner, 1931). Thus the prediction of the judgment involves differential weighting for the influence of the standard and the various features of the stimulus context (Helson, 1947; Helson, Michels, & Sturgeon, 1954).

Absolute judgments are distinguished by the absence of any specified standard so that the scale of judgment is more completely dependent upon the context stimuli. Consequently, the role of the various features of the stimulus context can be more readily determined.

The present studies were designed to further our understanding of context effects by systematic manipulation of the context stimuli for absolute judgments. The first experiment was planned to test *adaptation level* theory as elaborated by Helson (1947) and Johnson (1955). While the theory yields some relatively unambiguous deductions, there have been no studies which manipulated the stimulus context over a sufficient array of values to provide a clear basis for evaluation and improvement. Perhaps the most adequate tests have been made by Johnson (1944, 1949), who used the method of single stimuli to study the effects of different distributions of weights and tones. Johnson's results were consistent with his theory, but the number of different distributions

[1] Research supported in part by National Science Foundation Grant G-7485. Special thanks are due Norman H. Anderson for a critical reading of the manuscript.

of stimuli and also the number of *Ss* judging each of his distributions were too small (relative to the variability among *Ss*) to provide a sensitive test.

The initial objective of the present research was to obtain judgments of a much greater array of stimulus distributions than had ever been studied before. Both the method of presentation and the stimulus dimension were selected with a view to simplifying the relationships between stimuli and judgments. All of the relevant stimuli were presented simultaneously to maximize *Ss'* information about the stimulus values. An abstract dimension, numerical magnitude, was chosen to simplify the perceptual relationships among the stimuli.

Method

The features which were common to all four experiments will be described in this section; those particular to the separate experiments will be given in the appropriate sections.

Stimuli.—The stimulus context for each *S* consisted of a set of approximately 44 numerals, mimeographed on a single 8½ × 11-in. page. The instructions for judgment were printed at the top of the page, and the numerals were arranged in four columns beneath the instructions. Except when otherwise noted, the set ranged from 108 to 992, with no repetitions. The standard arrangement was in order of increasing magnitude down the columns, with 108 at the top of the first (left margin) column and 992 at the bottom of the fourth (right margin) column. Table 1 characterizes the stimulus distributions with respect to the frequencies with which numerals from different parts of the range were presented. The numerals were approximately evenly spaced within each of the nine subranges, but tendencies to use exactly equal increments or to favor any particular third digit were avoided.

Subjects.—A total of approximately 450 junior high school students, and 900 introductory psychology students (freshmen and sophomores at Swarthmore College and the University of California, Los Angeles), formed the two samples used in these experiments. The selection of *Ss* was dictated by considerations of convenience; and while the data will be treated separately for these two samples, the research was not designed to analyze the observed differences between the samples. Any *S* who did not maintain an ordinal scale (i.e., order his categories in conformity with the order of the numerals on the page) and did not use all of the available categories was eliminated under the assumption that he had not understood the instructions. While this involved considerable selection for the younger *Ss*, the selection did not appear to be systematically related to the experimental conditions. Each *S* judged only one set of numerals, with between 15 and 51 *Ss* judging each set.

Procedure.—The stimulus sheets were distributed in the classrooms, *Ss* in neighboring seats always receiving different sets of numerals. The printed instructions (also read aloud to the younger sample) were as follows:

This is a study of how people judge the size of numbers. On the bottom of the page, there is a list of numbers placed in the order of their size from very small to very large. Your job is to study the entire list and then to decide how large or small each number is—in comparison with all the others on this page. Write down on the dotted line next to each number how large it appears to you. For each number, choose one of the following sizes: very large—VL; large—L; medium—M; small—S; very small—VS. Thus if a number seems to be of average or medium size (compared with the others), write the letter M next to it; if you judge it to be very small, write VS. Be sure to write one of the five size abbreviations (VL, L, M, S, VS) next to each number. Take your time with these judgments, but do not worry about tricks or fancy arithmetic. The purpose is to find out how you compare the sizes of different numbers.

After several minutes, *Ss* were told to turn their papers over as soon as they were satisfied that they had correctly followed the instructions. The papers were collected after 10 min., most *Ss* requiring about half that time.

Analysis of the judgments.—Adaptation level theory has been primarily concerned with the prediction of the midpoint of the judgment scale, the adaptation level (AL). This is the value of the stimulus eliciting the central category of judgment (e.g., medium). When a range of stimuli elicit the central

category, some method must be employed to estimate the stimulus value within this range which would be judged central if *S* had to make a further subdivision. While presentations of the theory have not specifically dealt with this problem, the theory asserts that the AL for these "medium" stimuli would be a linear function of their mean (see below). Consequently, each *S*'s AL was defined as the arithmetic mean of the stimuli that he had judged medium. This value served as the dependent variable in all of the experiments, and the statistical evaluations of the effects of the different independent variables were based upon analyses of variance for these ALs.

Experiment I: The Mean

The fundamental assumption made by both Helson (1947) and Johnson

TABLE 1

Number of Stimuli in Respective Portions of the Range

Cond.	108–199	200–299	300–399	400–499	500–599	600–699	700–799	800–899	900–992
Experiment I									
A	12	13	11	3	4	0	0	0	1
B	12	13	10	1	2	1	2	1	2
C	8	8	8	8	8	1	1	1	1
D	7	6	7	8	6	2	2	2	3
E	10	6	3	2	2	2	3	6	10
F	1	2	4	9	12	9	4	2	1
G	5	5	5	5	5	5	5	5	5
H	3	2	2	2	6	8	7	6	7
I	1	1	1	1	8	8	8	8	8
Experiment II									
NG	3	4	4	6	9	12	6	0	0
NA	2	3	3	6	9	13	8	0	0
PA	0	0	8	13	9	6	3	3	2
PG	0	0	16	12	8	2	2	2	2
PGE	0	0	17	10	7	2	3	3	2
Experiment III									
M_1	16	16	12	14	16	9	7	6	4
M_2	20	20	18	10	6	4	4	7	11
M_3	31	21	4	6	5	5	5	7	16
M_4	4	6	7	9	15	15	12	16	16
M_5	11	7	4	4	6	10	18	20	20
M_6	16	7	5	5	5	6	4	21	31
LL	6	6	5	5	6	5	5	4	2
LH	9	6	3	2	2	10	8	2	2
HL	2	2	8	10	2	2	3	6	9
HH	2	4	5	5	6	5	5	6	6
PM	8	5	4	3	7	6	4	4	3
QM	3	4	4	6	7	3	4	5	8
RM	16	2	1	2	12	6	1	2	2
Experiment IV									
N	7	7	7	6	6	7	4	0	0
P	0	0	4	7	6	6	7	7	7
J	12	13	11	3	4	0	0	0	0
S	0	0	0	0	0	0	7	11	26

(1955) is that each stimulus pulls the AL, the midpoint of the judgment scale, toward its own perceptual value. The resultant AL is assumed to be a linear function of the arithmetic mean of the perceived values of the stimuli presented for judgment. However, since the perceived values are not directly known, it is necessary to rescale the physical values so as to conform to current hypotheses about the perceptual effects of the stimuli. In practice, both Helson and Johnson have gone along with Fechner by taking geometric rather than arithmetic means of the physical stimuli. An advantage of numerical magnitude, the stimulus dimension studied in the present experiments, is that no rescaling would appear necessary since discrimination should be of equal difficulty over the entire range of numerals. The hypothesis, then, is that AL is linearly related to the arithmetic mean of the set of numerals presented for judgment. This is tested in Exp. I by presenting different sets of numerals for judgment and observing the relationship between the resulting ALs and the arithmetic means of the different sets.

Method

The junior high *Ss* were presented the nine distributions described in Table 1. It should be noted that the selection includes distributions of varying degrees of both positive and negative skewness as well as normal, U-shaped, and rectangular distributions. In addition, five of these distributions were presented in reversed order, with 992 at the top left, and the other numerals decreasing in value down to 108 at the bottom right. These five reversed-order distributions (one of which was also presented in random order) are identified in Table 2 by the addition of an R to the condition label.

The college *Ss* judged only the distributions indicated in Table 3. These were all presented in the regular order, small to large. The number of *Ss* judging each distribution, junior high and college, is presented in Tables 2 and 3, respectively.

TABLE 2

JUNIOR HIGH SCHOOL ADAPTATION LEVELS: EXP. I

Cond.[a]	*N*	Mean AL	*SD*	Cond.[a]	*N*	Mean AL	*SD*
A	20	308	39	F	23	496	45
B	21	296	79	FR	21	537	40
C	29	368	83	G	25	489	104
D	27	403	62	GR	29	553	94
DR	29	467	74	H	23	612	62
DR_a	24	435	79	HR	24	654	71
E	25	464	133	I	24	624	73
ER	24	492	104				

[a] Addition of R to label signifies that the distribution identified by the first letter was presented in reversed order; DR_a was presented in random order.

Results

Junior high ALs.—The means and *SDs* of the ALs for the younger sample are presented in Table 2. The ALs for the nine of these groups which were presented the numerals in order of increasing magnitude are plotted against the stimulus means in Fig. 1. A test for deviations from linearity of regression yielded an $F = 1.18$, $df = 7/208$, indicating that the least-squares line (slope = .85, origin = 27) gives an adequate fit to

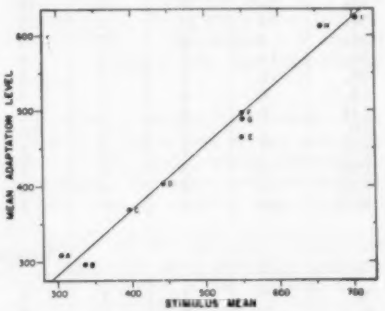

FIG. 1. Relationship between AL and stimulus mean for junior high groups judging numerals in ascending order.

TABLE 3

STIMULUS PARAMETERS AND COLLEGE ADAPTATION LEVELS

Cond.	N	Stimulus Parameters				Adaptation Levels			
		Mean	Midpoint	Median	Range	AL_p[a]	AL_∞	SD	% Error[b]
Experiment I									
A	32	304	550	277	900	401	332	68	7.7
C	16	396	550	378	900	446	430	62	1.8
D	16	443	550	419	900	465	465	40	0
F	18	550	550	551	900	524	539	18	1.7
G	15	551	550	552	900	525	545	29	2.2
I	17	704	550	723	900	602	619	39	1.9
Experiment II									
NG	30	517	425	558	650	466	481	44	2.3
NA	30	551	425	593	650	482	494	70	1.8
GC	37	551	550	552	900	525	532	40	.8
PA	31	551	675	509	650	581	567	61	2.2
PG	28	497	675	433	650	546	553	68	1.1
PGE	30	513	675	451	650	554	536	81	2.8
Experiment III									
M_1	33	449	550	450	900	479	494	51	1.7
M_2	34	444	550	350	900	434	437	91	.3
M_3	27	449	550	250	900	389	462	106	8.1
M_4	51	651	550	650	900	569	561	81	.9
M_5	46	654	550	750	900	614	610	92	.4
M_6	27	651	550	850	900	659	619	100	4.4
LL	20	500	550	500	900	501	494	38	.8
LH	21	500	550	600	900	546	548	57	.2
HL	21	600	550	500	900	501	487	44	1.6
HH	20	600	550	600	900	546	542	55	.4
PM	33	505	550	550	900	524	532	39	.9
QM	33	595	550	550	900	524	551	41	3.0
RM	33	433	550	550	900	524	527	38	.3
Experiment IV									
N	30	425	425	423	650	405	419	24	2.2
P	30	674	675	677	650	656	650	55	.9
J	32	288	350	272	500	300	287	30	2.6
S	32	901	850	925	300	873	891	25	6.0

[a] AL predicted from the following regression equation: $AL_p = .547(\text{Midpoint}) + .450(\text{Median}) - .027(\text{Range})$.

[b] Difference between obtained and predicted ALs divided by .01(Range).

the data. This result is consistent with the prediction from the Helson and Johnson theory of adaptation level.

Two additional groups were tested on U-shaped distributions similar to Cond. E but with very different means. For Group 10E ($N = 31$), each of the numerals in Cond. E was multiplied by 10 so that the new distribution ran from 1085 to 9925. For Group 0.1E ($N = 29$), each of

the numerals in Cond. E was multiplied by 0.1 and rounded to a whole number so that the new distribution ran from 11 to 99. The obtained ALs for these two groups were 4992 and 51. These values might have been predicted by multiplying the least-squares prediction for E by 10 and 0.1, respectively (predicted ALs = 4945 and 49). A corresponding Group D/10 (N = 23) yielded an AL = 38 as compared with a predicted AL = 40 (1/10 the least-squares prediction for D). It would seem that multiplying each of the stimulus values by a constant factor changes AL by the same factor. The rationale for this, in terms of adaptation level theory, is not immediately clear. It depends upon the interpretation of the constant term in the least-squares line, the AL intercept. If this value represented the effects of either past experience or background stimuli not being manipulated by E, some *ad hoc* assumptions would have to be made to explain why it should change by the same factor as the stimuli presented for judgment. This problem will be taken up again in the discussion following Exp. IV.

College ALs.—The ALs for the college *Ss* are presented in Table 3 along with the ALs for the conditions which will be described in later sections. The least-squares slope relating AL to stimulus mean is smaller than for the junior high groups, and the ALs are larger.

Ordinal effects.—The ALs for the five junior high groups presented the numerals in both forward and reversed orders were analyzed as a 5 × 2 design with *Ss* randomly eliminated so that each pair had equal N. Both ordinal and distribution effects were highly significant: $F = 22.3$, $df = 1/130$; $F = 36.4$, $df = 4/130$. However, the interaction F was less than one. The ALs for the five reversed-order conditions have a least-squares slope of .87 when plotted against the arithmetic means. This is approximately the same as the slope for the forward-order groups; however, the value of the AL intercept is raised by 33. Since the ALs for judgments of length (which also showed no interaction with distribution; Parducci, 1959b) shifted toward the stimulus values presented and judged first, one might assume here that *Ss* tend both to look at the numerals and subsequently to record their judgments in the order of appearance upon the page. The fact that the AL of the junior high, random-order group, DR_s in Table 2, was intermediate in value between the ALs for the junior high D and DR groups is also consistent with this interpretation. However, a college Group DR (N = 25) yielded an AL of 459 (SD = 49). This was actually lower, although not significantly so, than the AL for the college Group D, suggesting that any ordinal effects for college *Ss* were not determined by the order of the numerals on the page (i.e., these *Ss* might judge numerals in order from very small to very large regardless of the order on the page). That college *Ss* also show ordinal effects is suggested by the category-width analyses described next.

Category widths.—For each S, a tabulation was made of the range of numerals included in each of his five respective categories of judgment. This was obtained by subtracting the mean of the two numerals bounding the lower limen of each category (arbitrarily taken as 100 for the "very small") from the mean of the two numerals bounding the upper limen of the category (taken as 1000 for the "very large"). Mean category widths for four of the conditions are graphed

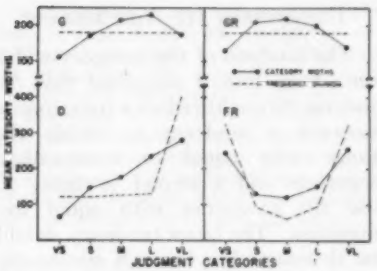

FIG. 2. Mean category widths for junior high groups. Dashed lines are category widths which would be obtained if the respective categories were used with equal frequency.

in Fig. 2. These four were selected to illustrate major trends found in the category-width data.

Comparison of Cond. G and GR shows the characteristic shift associated with the order of the numerals upon the page. Although there is a general tendency to use the middle categories more frequently than the two end categories, it is greatly modified in G by the relative narrowing of the "small" as compared with the "large" categories. This measure of the ordinal effects is quite large relative to the total range and is statistically significant when the analysis is based on the difference in category widths ("large" minus "small") for the two orders of the D, G, and H distributions. It is interesting that the GR curve is not a mirror image of the G curve and that none of the other pairs form mirror images, either. Analogy with the previous research on ordinal effects for judgments of visual length (Parducci, 1959b) suggests that under the reversed-order conditions, many of the junior high *Ss* work through the numerals in order from small to large, rather than from upper left to lower right on the page. Thus the reversed-order curves appear to represent judgment scales which are freer of systematic ordinal effects. The category-width curves for the college groups which judged the symmetrical distributions (i.e., E, F, and G—not illustrated) also have the asymmetry shown for G, suggesting that they also reflect some kind of ordinal effect.

A second major trend in the data, illustrated by the D and FR graphs, is the tendency to compromise between equal-width and equal-frequency divisions of the stimulus sets. If *Ss* had divided the stimulus range (approximately 900) into five equal parts, the category-width curves would all have been horizontal at 180. If they had divided the 40-odd numerals up into five subgroups with the same number of numerals in each, the respective category-width curves would have been those shown by the broken lines. For Group FR, the mean category widths are in each case intermediate between these two possible values. This can also be seen (though somewhat less clearly) in the D graph where the "very large" category is much wider than in the G graph but yet not wide enough to include one-fifth of the numerals.

EXPERIMENT II: THE MIDPOINT OF THE STIMULUS EXTREMES

While the correlation between AL and stimulus mean was almost perfect in Exp. I, the end values were the same for all stimulus sets (108 and 992). There was thus no opportunity to determine whether the stimulus extremes, or their own mean, hereafter called the midpoint, have any *special* effect upon AL, apart from their contribution to the mean. Adaptation level theory has placed on special weighting upon the end values, but they have been stressed in some

other approaches to judgment (cf., Volkmann, 1951).

Method

To assess the effects of these end values, five new groups of college students were exposed to the distributions of stimuli shown in Table 1. A sixth group, GC, judged the rectangular distribution, G, already judged by junior high and college groups in Exp. I. The NA and PA distributions were constructed so that they had the same arithmetic means as distribution G, but their end values were varied to make 749 the largest numeral in the NA set and 351 the smallest in the PA set. The NG and PG distributions had the same end values as NA and PA, respectively, but their geometric means were made equal to the geometric mean of the rectangular distribution, 476. These two groups were run to determine whether the difference between the NA and PA geometric means, 516 and 530 respectively, was responsible for any observed differences in AL. The PGE distribution was the result of an error in selecting the numerals for a PG distribution. It is thus slightly different from the PG both in respect to its geometric and arithmetic means. It should also be noted that the medians varied systematically for these six groups, the N distributions having higher medians than the P distributions. The medians and also the midpoints (i.e., the arithmetic means of the two end values) are presented in Table 3.

Results

The ALs for these six groups, presented in Table 3, differ significantly ($F=8.66$, $df=5/180$, $P<.001$). It should be noted that the three groups with the highest midpoint all have higher ALs than the two with the lowest midpoint, while the AL of the rectangular group is of intermediate value. These data thus indicate that the mean of the stimuli (arithmetic or geometric) is not the sole determiner of AL; and they are consistent with the hypothesis that the extreme values of the stimulus distribution have a special effect upon AL, pulling it toward their own mean, the midpoint.

Experiment III: The Median

The analysis of the category-width data from Exp. I suggested that *Ss'* judgments might reflect a compromise between a tendency to divide the range into equal or symmetrical segments and a second tendency to use the categories with equal frequencies. The latter tendency should be demonstrable through systematic manipulation of the median of the stimulus sets presented for judgment since the higher the median the greater the range of the numerals which would be included in the categories below AL. The median was positively correlated with the mean in Exp. I and negatively correlated with the midpoint in Exp. II. The distributions studied in Exp. III were designed so that the median varied independently of the mean and midpoint. If there is the hypothesized tendency to use the categories above and below AL with equal frequency, then AL will vary directly with the median even though the mean and midpoint are constant.

Method

For the purpose of simplifying the exposition, the description of the new conditions will be subdivided into three subexperiments: A, B, and C.

Experiment A.—Six groups of college *Ss*, M_1 through M_6, were exposed to the distributions characterized in Tables 1 and 3. The midpoint is 550 for all six groups, and the mean is approximately 450 (for M_1, M_2, and M_3) or 650 (M_4, M_5, and M_6). In order to obtain such large variations in median without introducing large gaps between successive numerals, each of the six M distributions contained 100 instead of the usual 44 numerals. A further variation was the use of four rather than five categories of judgment, the "medium" category being eliminated.

Experiment B.—Four conditions (LL, LH, HL, and HH) were used in a 2×2 factorial design in which mean and median were each varied, with the midpoint constant at 550 under all four conditions. The magni-

tude of variation in median was smaller here than in Exp. A, and the conditions corresponded more closely to those of the previous experiments (44 numerals under each condition with five judgment categories).

Experiment C.—Three conditions, PM, QM, and RM, had *different* means but identical medians and end values. These served to supplement that portion of the analysis of the Exp. B results which was concerned with the effects of variation in the mean.

Results

The ALs, shown in Table 3, were calculated as before except that the mean of the smallest numeral judged "large" and the largest numeral judged "small" was taken as the AL for the six groups of Exp. A. The variation in median significantly affected the ALs of the three high-mean groups in Exp. A ($F = 8.01$, $df = 2/121$, $P < .001$) and also the ALs for the corresponding low-mean groups ($F = 3.69$, $df = 2/91$, $P < .05$). However, inspection of these ALs (in Table 3) suggests that a point of diminishing returns for this effect appears to have been reached by the two extreme groups, M_3 and M_6. The difference between the ALs for M_2 and M_3, which is in the direction opposite to that of the predicted difference, does not approach statistical significance.

The results of Exp. B also indicate that the median determines AL. While neither the variation in stimulus mean nor the interaction between mean and median approached significance ($F < 1.0$ in both cases), the median again has marked effects with mean held constant ($F = 23.2$, $df = 1/76$).

The results of Exp. C, in which the mean varied but the median and midpoint did not, are consistent with Exp. B in showing relatively small variation in AL associated with the rather large variation in mean. However, here the ALs are positively correlated with the means, and the differences in AL are statistically significant at the .05 level ($F = 3.43$, $df = 2/96$).

Experiment IV: The Midpoint and the Median

Method

The four additional conditions in Tables 2 and 3 are treated together here because they provide an additional basis for describing AL as a function of the alternative central tendency values of the stimulus distribution. The J and S distributions were drawn up to provide a greater variation in midpoint than had been investigated in Exp. II. The N and P distributions are rectangular with midpoints corresponding to the two extremes of Exp. II. Since the medians and means vary along with the midpoints of these four new conditions, the ALs are of theoretical interest only when considered along with the ALs for the various other combinations of these measures described in the preceding experiments.

Results

The four new ALs, together with the 25 other college ALs included in Table 3, constitute the data treated in this section. These 29 mean ALs were used in a multiple regression analysis with various combinations of the midpoint, median, range, and mean as predictor variables. The results of the analysis are presented in Table 4. This multiple regression technique permits the separate assessment of the criterion variation (i.e., of AL) accounted for by any subset of predictors, and of the additional criterion variation accounted for when the remaining predictors are also used (Quenouille, 1952). Thus the second line of the table indicates that addition of the mean to the other three predictors does not significantly increase the proportion of the variation in AL predicted by the multiple regression based on the three predic-

TABLE 4

ANALYSIS OF VARIANCE FOR ADAPTATION LEVELS: BASED UPON MULTIPLE REGRESSIONS

Source	*df*	*F*	*P*
Due to midpoint, median, & range	3	1575.704	<.001
Extra due to mean	1	3.786	>.05
Due to midpoint, range, & mean	3	2098.660	<.001
Extra due to median	1	6.835	<.05
Due to median, range, & mean	3	2090.943	<.001
Extra due to midpoint	1	29.987	<.001
Due to midpoint & median	2	3149.376	<.001
Extra due to range & mean	2	1.994	>.05
Unaccountable (Error)	25	(1309.684)	
Total	29		

tors in the first line of the table. The general conclusion of the analysis is that the mean and range make no significant contribution to the predictive power of the median and midpoint combined. However, failure to include either the midpoint or the median would result in a significant loss of predictive power.

The predicted values of AL, AL_p in Table 3, were based upon the following regression equation:

$$AL_p = .547(X_1 + X_n)/2 + .450(X_{(n+1)/2}) - .027(X_n - X_1)$$

where X_1 and X_n are the smallest and largest numerals in the distribution of stimuli presented for judgment. The first term on the right is the midpoint, the second the median, and the third the range. Although the contribution of the range was not statistically significant, it has been included for theoretical reasons which are presented below in the discussion of ordinal effects.

DISCUSSION

The Mean

Experiment I was designed to investigate implications of adaptation level theory in a simple experimental situation. For nine different sets of stimulus values, the ALs clearly appeared to be a linear function of the stimulus means. This constitutes perhaps the most extensive support adaptation level theory has had concerning the specific effects of unweighted context stimuli. However, it was also found that multiplying each stimulus in the distribution by a constant factor had the effect of multiplying AL by the same factor. This did not appear to follow from adaptation level theory. Furthermore, manipulation of the mean, with midpoint and median held constant, did not have a significant effect on AL in Exp. III-B (AL actually varying inversely with the mean), and only a slight effect in Exp. III-C. Finally, the multiple regression analysis, using all 29 college ALs, indicated that the mean did not add significantly to the accuracy of the prediction based upon the other stimulus parameters. Thus the mean does not appear to provide an adequate basis for description of the results of these experiments.

Ordinal effects.—The differences in AL (and category limens) associated with the order of the numerals on the page in Exp. I appear analogous to the ordinal effects previously reported for judgments of visual length. The experimental analysis of the ordinal effects for length indicated that adaptation level theory would need further elaboration before it could handle the special tendency to switch categories with discriminable changes in the stimulus values (Parducci, 1959b). Since the magnitude of the successive changes in stimulus values would on the average be proportional to the range, the range

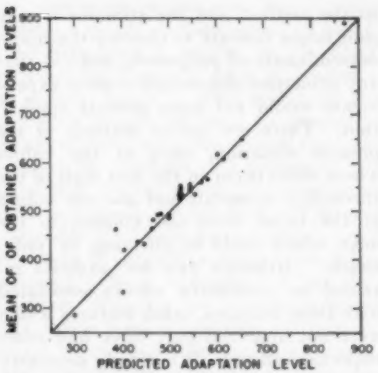

FIG. 3. Relationship between obtained and predicted ALs, with prediction based on regression coefficients of .547, .450, and −.027 for midpoint, median, and range, respectively.

was included as a third variable in the multiple regression equation for the predictions graphed in Fig. 3. However, the effects of order were relatively small for the college *S*s and could actually have been neglected in the analysis.

The Midpoint and Median

The location of the respective category limens in Exp. I suggested that *S*s' judgments may be represented as a compromise between two, sometimes contradictory, tendencies. One of these tendencies or habits fixes the proportionate widths of the respective categories of judgment. Thus if *S* were permitted only two categories, he would tend to divide them at the midpoint of the stimulus distribution. His second tendency, which fixes the proportionate frequencies with which the respective categories are used, would be to make the division at the median of the stimulus distribution. However, the *S*s in the present experiments (except those in Exp. III-A) used five categories so that the midpoint of the judgment scale was always some hypothetical division of the "medium" stimuli. In conformance with adaptation level theory, this division was assumed to be at the mean of the stimuli judged "medium." The assumption suggested by the present analysis is that the actual division would be determined by a compromise between their midpoint and median. Since the mean of the "medium" stimuli would be a close approximation to the midpoint-median compromise for these same stimuli, their mean was used as the measure of AL in all of the analyses.

The results of Exp. II and III indicated that systematic variation of either the midpoint or the median produces a corresponding variation in AL, even though the mean was kept constant. Exclusion of either variable from the multiple regression analysis (for all 29 ALs) resulted in a significant loss of predictive power. The theory of adaptation level would have to be modified to handle these results. They appear instead to follow from the operation of the proportionate-width and -frequency tendencies. The fact that the regression coefficients for the midpoint and median total so closely to 1.0 (.997) is consistent with the assumption that the judgment process is basically a compromise between these two tendencies.

Multiplying each stimulus in the distribution by a constant factor (Exp. I) had the effect of multiplying AL by the same factor. This also is consistent with the midpoint-median hypothesis. Since both the midpoint and median would be multiplied by this factor, the same would be true for the AL predicted from a linear combination of the midpoint and median.

Negative instances.—Inspection of Table 3 reveals that the predicted ALs are most in error for the A and M_3 distributions. The error of the midpoint-median prediction for A seems related to the fact that this was the only stimulus series containing a gap greater than 100 between successive numerals (the gap was from 592 to 992). A number of experiments have indicated that beyond a certain critical value, further increases in the difference in magnitude between an extreme or anchoring stimulus and the other stimuli being judged have diminishing effects upon the judgment scale (e.g., Postman & Miller, 1945;

Rogers, 1941; Sherif, Taub, & Hovland, 1958).

The relatively large error of the prediction for M_3 points up a limitation in the use of the median as a predictor of AL. In both this distribution, and also $M_{6'}$ there were only two numerals (out of a total of 100) separating the obtained from the predicted ALs. Since the errors in prediction, 73 and 40 respectively, were both in the direction of reduced difference between AL and the midpoint, an extremely small, but systematic, shift in the relative frequency with which respective categories were used permitted sizable reductions in the differences between the widths of the categories above and below AL. This suggests that the compromise might involve heavier weighting of the midpoint whenever the median and midpoint are very different in value but are separated by only a very few stimuli. It should be noted, however, that M_3 and M_6 are extremely irregular distributions, their medians being separated from the nearest end value by only one-sixth of the range.

Implications of the midpoint-median modification of AL theory.—Perhaps the most critical concern with the results of this research should be with the nature of the stimulus materials and the method by which they were presented for judgment. Numerical magnitude, as represented by numerals arranged on a page, is an unconventional stimulus dimension, considerably different in character from the lifted weights and pure tones of classical psychophysics. Should adaptation level theory be expected to apply to the present situation? Neither Helson nor Johnson have restricted the realm of applicability of their approach to judgment, and the experimental assumption here has been that principles of judgment would manifest themselves more readily with such an abstract dimension, unclouded by complex "receptor functions," and presented so as to permit *Ss* to grasp as easily as possible the relationships among the different stimuli forming the context for judgment.

The dangers of this procedure would appear to be: (*a*) that features peculiar to the method and the stimulus dimension might operate to obscure the major determinants of judgment, and (*b*) that any principles discovered in these experiments would not have general application. There are special features of the present situation, such as the values where shifts occur in the first digit of the three-digit numerals and also the values at the break from one column to the next, which could be affecting *Ss*' judgments. Although *post hoc* analyses revealed no systematic effects associated with these features, other features which were not analyzed may have had more important effects. It is clearly necessary to conduct similar research using different kinds of stimuli and different methods of judgment. Results similar to those found here have recently been reported for perceptual judgments of the numerousness of different collections of dots presented by the method of single stimuli, AL being found to vary with the midpoint and median even though mean was held constant (Parducci, 1959a). Research is in progress studying these context effects upon judgments of visual size and length and also of lifted weights.

The advantages of using a combination of the midpoint and median, instead of the mean, would be greatly enhanced if the conditions affecting their relative weighting could be determined. Presumably, the weighting of the median would be reduced if *S* did not have a continuous record of his own response frequencies before him. Thus for any stimulus dimension, the weighting should shift depending upon whether *S* keeps the record of his judgments or *E* records them privately. The weighting might also be expected to shift depending upon the difficulty of remembering the end values of the stimulus context. With stimulus dimensions for which these values would be particularly difficult to discriminate (e.g., lifted weights), the median should have relatively greater influence.

Insofar as comparative judgments are also affected by the stimulus context, the midpoint-median compromise should

be operating—sometimes to produce constant errors of judgment. Thus if there has been a surplus of "smaller" judgments, *S* should have a tendency to report "larger." This might be the case even though the mean (or geometric mean) of the comparison stimuli has been equal to the standard. And similarly, the particular values of the extreme comparison stimuli, the smallest and largest variables, may have greater effects upon the point of subjective equality than has commonly been assumed. These hypotheses, of course, are predicated on the adaptation level explanation of constant errors in terms of the stimulus context. The modification suggested here is concerned solely with those features of the stimulus context which are most useful for the prediction of judgment.

While the major results of the present experiments suggest that the predictive value of the mean depends on its relationship to the midpoint and median, it should not be inferred that the mean is always a poor first-approximation to AL. In fact, if the proportionate-width and -frequency interpretation is correct, the mean will be a good predictor whenever it falls between the midpoint and median. Since this is the case for most distributions of stimuli (e.g., all those chosen for Exp. I), it is possible that the success of the mean has obscured the role of these other two variables.

Summary

The effects of different stimulus contexts were studied by presenting different sets of numerals for judgment in terms of an absolute scale of numerical magnitude. The major independent variables were the mean, the midpoint (mean of the two end values), and the median of the stimulus context. Contrary to the theory of adaptation level, it was found that shifts in judgment were associated with shifts in either the midpoint or the median—even though the mean was held constant. The mean itself appeared to have little effect upon judgment. The data were interpreted as consistent with the proposal that the judgment scale reflects a compromise between two different tendencies: (*a*) to divide the range into proportionate subranges, and (*b*) to use the alternative categories of judgment with proportionate frequencies. The success of the mean as a rough predictor of adaptation level may thus follow from the fact that for most distributions of stimuli, the mean is intermediate in value between the midpoint and median.

References

Garner, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48,** 218–224.

Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60,** 1–29.

Helson, H., Michels, W. C., & Sturgeon, A. The use of comparative rating scales for the evaluation of psychophysical data. *Amer. J. Psychol.*, 1954, **67,** 321–326.

Johnson, D. M. Generalization of a scale of values by the averaging of practice effects. *J. exp. Psychol.*, 1944, **34,** 425–436.

Johnson, D. M. Generalization of a reference scale for judging pitch. *J. exp. Psychol.*, 1949, **39,** 316–321.

Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

Künnapas, T. M. Influence of frame size on apparent length of a line. *J. exp. Psychol.*, 1955, **50,** 168–170.

Parducci, A. A proposed modification of Helson's theory of adaptation level. Paper read at West. Psychol. Ass., San Diego, April 1959. (a)

Parducci, A. Ordinal effects in judgment. *J. exp. Psychol.*, 1959, **58,** 239–246. (b)

Postman, L., & Miller, G. A. Anchoring of temporal judgments. *Amer. J. Psychol.*, 1945, **58,** 43–53.

Quenouille, M. H. *Associated measurements.* New York: Academic Press, 1952.

Rogers, S. The anchoring of absolute judgments. *Arch. Psychol.*, 1941, **37,** No. 261.

Sherif, M., Taub, D., & Hovland, C. I. Assimilation and contrast effects of anchoring stimuli on judgment. *J. exp. Psychol.*, 1958, **55,** 150–155.

Turner, W. D. Intra-serial effects with lifted weights. *Amer. J. Psychol.*, 1931, **43,** 1–25.

Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.

(Received August 22, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 78–82

STIMULUS DIFFERENCES IN DISCRIMINATION LEARNING

R. A. CHAMPION AND C. D. STANDISH

University of Sydney

The following study was suggested by an experiment (Bitterman, 1952) in which rats were trained to jump to the left member of a pair of cards showing wide stripes (on the left) and narrow stripes (on the right) and to jump to the right member of a pair consisting of a light grey card (left) and a dark grey card (right). When this group was tested in the usual nonspatial discrimination learning situation, including stimuli in the reversed spatial relations as well as the pairs described above, there was little or no positive transfer. This finding was interpreted by Bitterman as contrary to an implication of the "approach" theory (Nissen, 1950; Spence, 1936). It was argued that if the learning in the first stage had consisted of approach to the positive stimuli (wide stripes and dark grey) then the reversal of spatial relations in the test situation should have had no effect; that is, *S*s should have continued to approach the previously correct stimuli (continuity of learning and positive transfer). On the other hand the "configurational" theory (Gulliksen & Wolfle, 1938) was said to be supported because *S*s jumped left to narrow stripes and right to light grey when presented with the transposes in the second stage, implying that the perceived configurations had not changed with reversal.

The specific purpose of the present experiment was to test the possibility that between-pair and within-pair differences play an important role in the type of situation described above. In Bitterman's experiment the between-pair difference was deliberately made greater than the within-pair differences to allow for the possibility of "configurational perception" (Bitterman, 1952, p. 173). This case has been dealt with in terms of the "approach" theory as well as the "configurational" theory (Nissen, 1952) but the two types of explanation might also be contrasted in the case where the within-pair differences are greater than the between-pair difference.

METHOD

Subjects.—The *S*s were 60 students from an introductory course in psychology at the University of Sydney, divided randomly into four groups of 15. Colorblind students were excluded from the outset.

Apparatus.—A Pickford anomaloscope (Pickford, 1953) was used to present to *S* colored circles of light 1 in. in diameter and 3 in. apart in the horizontal midline of a 5 × 2-in. milk glass screen approximately 18 in. in front of *S*. The duration of the stimuli on any trial was held constant at 100 msec. by means of an electronic timer. The *S* was required to press one of two buttons on each trial as a response, the buttons being switches mounted in a box placed on the table below and to the right of the stimulus screen. The switches were placed 5½ in. apart in the 7 × 7-in. horizontal top surface of the box on the left and right of center corresponding to left and right stimuli in the vertical plane of the glass screen. The stimulus screen was surrounded by an opaque screen and the arrangement of the apparatus was such that *E* could observe *S*'s responses without being seen by *S*.

Stimuli.—The stimuli were chosen so as to allow two basic conditions: (*a*) between-pair difference greater than within-pair differences, and (*b*) within-pair differences greater than between-pair difference. There were four

TABLE 1

STIMULUS PAIRS USED FOR THE FOUR GROUPS OF *Ss* IN THE TWO STAGES OF THE EXPERIMENT[a]

Stage	Experimental Groups		Control Groups	
	1 $b > w$	3 $w > b$	2 $b > w$	4 $w > b$
1	$R_1+:R_3-$ $G_2-:G_1+$	$G_1+:R_3-$ $G_2-:R_1+$	—	—
2	$R_1+:R_3-$ $G_3-:G_1+$ $R_3-:R_1+$ $G_1+:G_3-$	$G_1+:R_3-$ $G_3-:R_1+$ $R_3-:G_1+$ $R_1+:G_3-$	$R_1+:R_2-$ $G_3-:G_1+$ $R_2-:R_1+$ $G_1+:G_2-$	$G_1+:R_2-$ $G_2-:R_1+$ $R_2-:G_1+$ $R_1+:G_2-$

[a] The spatial relations of the symbols in each pair correspond to the actual left-right arrangements of the stimuli. The designations $b > w$ etc. refer to relative magnitudes of stimulus differences between and within pairs.

stimuli with the following designations on the Munsell system according to visual matchings by *E*: 7.5 R 5/12 (R_1), 10 R 7/8 (R_2), 5 G 8/8 (G_1), and 5 Y 9/10 (G_2). Stimulus R_1 resembled R_2 and G_1 resembled G_2, but R_1 and R_2 differed considerably from G_1 and G_2.

Instructions to S.—Each *S* was given the following instructions:

In this experiment I want you to place the index finger of your right hand on the center mark between the two buttons. Every 30 sec. or so two lights will flash on the screen in front of you. When you see these two lights all you have to do is press one or other of the two buttons. As the lights will be visible only for one-tenth of a second, I will say 'ready' just before the lights are flashed. If you have pressed the correct button I will say 'correct'; if the incorrect button is pressed I will say 'wrong.' Press the button firmly down and keep your finger on it until I tell you whether you are correct or not. At first it will be difficult to know which is the correct button to press and you will just be guessing. Later you will learn which is the correct one and should have no difficulty in your selections. If you realize you have made an incorrect response or if I tell you that you are wrong, do not press the other button but return your finger to the central position as you would for a correct response. Have you any questions?

Procedure.—The procedure for each of the four groups of *Ss* was as follows. Group 1 was an experimental group trained with between-pair differences dominant. These *Ss* were first trained in Stage 1 with R_1 (positive) and R_3 (negative) and with G_2 (negative) and G_1 (positive) as the two pairs to be discriminated with constant spatial arrangement. They were then presented in Stage 2 with the transposes as well as the original pairs, i.e., with R_3 (negative) and R_1 (positive), and with G_1 (positive) and G_2 (negative) as well as the other two pairs. Group 2 served as a control for Group 1, learning only Stage 2, i.e., the usual nonspatial discrimination-learning situation. Group 3 was a second experimental group, trained with within-pair differences dominant. These *Ss* were first trained in Stage 1 with G_1 (positive) and R_3 (negative) and with G_2 (negative) and R_1 (positive), all in a constant spatial arrangement. They were then presented in Stage 2 with the transposes in addition, namely R_3 (negative) and G_1 (positive), and with R_1 (positive) and G_2 (negative). Group 4 was a control for Group 3, learning only the second of the arrangements presented to Group 3 (Stage 2). The order in which the stimulus symbols are listed above designates the left-right spatial positions of the lights in the apparatus from *Ss* viewpoint. The procedure and stimulus pairs are summarized in Table 1.

The *E* said "ready" approximately 2 sec. before flashing each stimulus pair on the screen. The order of presentation of pairs was randomized for each of the four groups with the restriction that no eight successive trials allowed simple right-left alternation among the correct responses. The criterion of acquisition for each *S* was eight successive correct trials.

RESULTS

The obtained mean scores for Stage 2 in terms of trials to criterion and errors are summarized in Table 2. Group 1 took longer to learn in Stage 2 than Group 2, displaying more errors and negative transfer. Group 3, however, learned more quickly in

TABLE 2

MEAN SCORES IN STAGE 2

Measure	Between-pair Differences Dominant		Within-pair Differences Dominant	
	Group 1 (Exp.)	Group 2 (Control)	Group 3 (Exp.)	Group 4 (Control)
Trials	39.1	24.6	19.0	28.0
Errors	16.8	9.1	4.9	9.2

Stage 2 than did Group 4 and made fewer errors, showing positive transfer. These data were subjected to analyses of variance in the expectation that the interaction between groups (experimental vs. control) and stimulus differences (between > within vs. within > between) would be significant (differential transfer). The results of the analysis of trials to criterion are presented in Table 3, in which it may be seen that the expectation is confirmed. The same result was obtained with an analysis of error scores. The expected form of interaction was confirmed in follow-up t tests ($df = 28$). Control Groups 2 and 4 did not differ significantly in their performance in Stage 2 (Table 2) either in terms of trials to criterion ($t = 0.46$, $P > .05$) or errors ($t = 0.10$, $P > .05$). A comparison of the experimental groups, however, showed that the prior learning in Stage 1 had a significant effect on performance in Stage 2 both in terms of trials to criterion ($t = 3.46$, $P < .01$) and errors ($t = 3.86$, $P < .01$).

A comparison of the performance of the two experimental groups (Groups 1 and 3) in Stage 1 is presented in Table 4, which shows differences favoring Group 1 but significant only with trials to criterion as a measure. The experimental groups learned both stages in succession and if the two tasks are treated as one by adding the trials to criterion then Group 3 (within-pair differences dominant) is superior to Group 1 (between-pair differences dominant), the means being 33.7 and 49.5 respectively ($t = 2.9$, $P < .01$ for 28 df).

TABLE 3

ANALYSIS OF VARIANCE IN TRIALS TO CRITERION IN STAGE 2

Source	df	MS	F
Exp. vs. Control Groups (A)	1	41.67	1.39
Type of stimulus difference (B)	1	516.27	17.17*
A × B	1	552.06	18.36*
Ss	56	30.07	

* $P < .01$.

TABLE 4

MEAN SCORES FOR THE TWO EXPERIMENTAL GROUPS IN STAGE 1

Measure	Group 1	Group 3	t
Trials	10.5	14.5	2.52*
Errors	1.5	3.1	1.18

* $P = .05$ when $t = 2.05$ for 28 df.

DISCUSSION

Despite the use of human *Ss* in a somewhat different type of situation, the results with the stimuli in which between-pair differences dominated (Groups 1 and 2) are similar to those obtained by Bitterman (1952). Under these conditions the "configurational" theory implies that the response would be primarily to "redness" with the paired Stimuli R_1:R_2 and to "greenness" with Stimuli G_2:G_1, so that the inclusion of the transposes requiring opposite responses in Stage 2 should conflict with the preliminary training in Stage 1 and produce negative transfer. With a strict interpretation of this theory the same reasoning should apply when within-pair differences dominate (Groups 3 and 4); once the discrimination between configurations G_1:R_2 and G_2:R_1 is made then the presentation of R_2:G_1 and R_1:G_2 requiring opposite responses should be equally disturbing. That this latter was not the case is clearly shown in the data. It might be suggested, however, that the latter configurations (R_2:G_1 and R_1:G_2) are new ones and therefore do not lead to negative transfer. If this were the case then the performance of the control groups (Groups 2 and 4) in Stage 2 might be expected to show a significant difference favoring Group 4. This is because Group 2 is

confronted with highly similar configurations demanding opposite responses whereas the members of Group 4, according to this version of the theory, simply respond to four different configurations from the outset. The difference obtained in this experiment favors Group 2, but it is not statistically significant. In general terms, if the "configurational" theory is to be interpreted in this way then a serious defect lies in its failure to specify before the event the conditions under which perceived configurations are identical and different.

Somewhat allied to the "configurational" formulation in the history of theories of discrimination learning is the "relational" theory (Lashley, 1942). Whereas the former theory would place emphasis upon differences between pairs of stimuli the latter theory refers rather to relations between stimuli *within* pairs (Birch & Vandenberg, 1955). This theory appears to break down when applied to the performance of Group 1 (between-pair difference dominant) where it would seem to lead to the expectation of positive transfer if *S* learns to respond to the redder or greener stimulus, the darker stimulus, or the more saturated stimulus, given in all cases by R_1 and G_1.

The "approach" theory apparently applies equally well to the performance of Groups 1 and 3 and its application is summarized symbolically in Table 5.

TABLE 5

STIMULUS COMPOUNDS ELICITING A CORRECT RESPONSE ACCORDING TO THE "APPROACH" THEORY[a]

Stage	Experimental Groups	
	Group 1 (Between > Within)	Group 3 (Within > Between)
1	$\left.\begin{matrix} S_{red} \\ S_L \end{matrix}\right\}$ $\left.\begin{matrix} S_{green} \\ S_R \end{matrix}\right\}$	$\left.\begin{matrix} S_{G1} \\ S_L \end{matrix}\right\}$ $\left.\begin{matrix} S_{R1} \\ S_R \end{matrix}\right\}$
2	$\left.\begin{matrix} S_{R1} \\ S_L \end{matrix}\right\}$ $\left.\begin{matrix} S_{G1} \\ S_R \end{matrix}\right\}$ $\left.\begin{matrix} S_{R1} \\ S_R \end{matrix}\right\}$ $\left.\begin{matrix} S_{G1} \\ S_L \end{matrix}\right\}$	$\left.\begin{matrix} S_{G1} \\ S_L \end{matrix}\right\}$ $\left.\begin{matrix} S_{R1} \\ S_R \end{matrix}\right\}$ $\left.\begin{matrix} S_{G1} \\ S_R \end{matrix}\right\}$ $\left.\begin{matrix} S_{R1} \\ S_L \end{matrix}\right\}$

[a] The subscripts R and L stand for cues on the right and left, respectively.

With the between-pair differences dominant (Group 1) *S*s in Stage 1 would presumably respond positively to the stimulus compounds red-left and green-right. Negative transfer occurs with the shift to Stage 2 because of a tendency also to respond to the compounds red (R_2)-left and green (G_2)-right, deriving in turn from stimulus generalization between R_1 and R_2 and between G_1 and G_2. When within-pair differences dominate (Group 3), however, *S*s in Stage 1 must presumably learn responses to the compounds G_1-left and R_1-right and must reduce generalization from these two compounds to G_2-left and R_2-right, respectively. In other words, *S*s must now respond differentially to R_1 and R_2 and to G_1 and G_2 as well as merely to red and green. If it now be granted that, as the "approach" theory implies, the two tasks presented to Groups 1 and 3 in Stage 2 are identical in requiring responses to G_1-left, R_1-right, G_1-right, and R_1-left (Table 5) then it follows that there should be positive transfer in Group 3, half these responses having already been learned in Stage 1. In keeping with these latter interpretations on the basis of the "approach" theory are the facts that Stage 1 was more difficult to learn with within-pair differences dominant (Group 3, Table 4) and that the two control groups did not differ significantly in performance in Stage 2 (Groups 2 and 4, Table 2).

Constant reference has been made to "stimulus compounds" in the application of the "approach" theory given above. No mention was made of these compounds or patterns in the first formulation of the theory (Spence, 1936) which was limited to "simultaneous" discrimination learning. The concept was applied later (Spence, 1952) to account for the case of "successive" discrimination learning in which no one stimulus is consistently reinforced more than any other member of the stimulus pairs. This restriction does not strictly apply in the present experiment but the process of patterning or compounding has nevertheless been assumed to occur

for two reasons. First, the similarity between stimuli in the pairs used produces a state of affairs similar to the restriction specified by Spence, if not identical with it. Secondly, Birch and Vandenberg (1955) have produced evidence that patterning may occur even under conditions of "simultaneous" discrimination learning when the restriction does not apply. It must be emphasized, however, that the stimulus patterns or compounds of S-R theory are to be clearly distinguished from perceived stimulus configurations or perceived relations between stimuli (Spence, 1952).

There has recently been proposed a "dual-process" theory of discrimination learning (Bitterman, Tyler, & Elam, 1955) which incorporates features formerly taken to characterize opposing theories, namely, learning to approach specific stimuli vs. learning to respond to total configurations. The proponents of the theory suggest that the former process accounts for "simultaneous" discrimination learning whereas the latter activity allows discriminations to be made in "successive" learning. The dual-process theory clearly applies to the data of the present experiment with the assumption that "configurational" learning occurs in Group 1 and "approach" learning in Group 3. For reasons of parsimony, however, it is suggested that the "approach" theory provides a better explanation of the present data.

Summary

An experiment was conducted to test the possibility that between- and within-pair differences play a significant part in the type of discrimination learning situation in which preliminary training is given with two pairs of stimuli in fixed spatial relations followed by test learning involving the two pairs and their transposes. In an earlier experiment with rats (Bitterman, 1952), in which between-pair differences in the stimuli were dominant, no positive transfer was apparent from preliminary to test learning. This finding was duplicated in the present study, with human *Ss*, but positive transfer was obtained under conditions in which within-pair differences dominated. These results were taken to be more in keeping with the "approach" theory of discrimination learning than with the "configurational" theory. Although a "dual-process" formulation applies to the findings of the present experiment, the "approach" theory is preferred as the more parsimonious.

References

Birch, D., & Vandenberg, V. The necessary conditions for cue-position patterning. *J. exp. Psychol.*, 1955, **50**, 391–396.

Bitterman, M. E. Approach and avoidance in discriminative learning. *Psychol. Rev.*, 1952, **59**, 172–175.

Bitterman, M. E., Tyler, D. W., & Elam, C. B. Simultaneous and successive discrimination under identical stimulating conditions. *Amer. J. Psychol.*, 1955, **68**, 237–248.

Gulliksen, H., & Wolfle, D. L. A theory of learning and transfer: I. *Psychometrika*, 1938, **3**, 127–149.

Lashley, K. S. An examination of the 'continuity theory' as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26**, 241–265.

Nissen, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.

Nissen, H. W. Approach and avoidance: A reply. *Psychol. Rev.*, 1952, **59**, 237–239.

Pickford, R. W. *Individual differences in colour vision*. London: Routledge and Kegan Paul, 1953.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Spence, K. W. The nature of the response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.

(Received July 23, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 83–88

ADDITIVITY OF CUES IN DISCRIMINATION LEARNING OF LETTER PATTERNS[1]

THOMAS R. TRABASSO

Michigan State University

In general, one says that cues are additive when it is found that *S*s can master a problem with two relevant cues (either of which can be used to solve the problem) more rapidly than a problem with only one relevant cue. The implication is that the two cues "add" in some manner. The theory under consideration here (Restle, 1955) makes explicit the underlying assumption that the rate of learning depends upon the number or strength of relevant cues, and, when formulated in set-theoretic mathematics (see below), states that it is the measures of sets of cues which may add. In finite set theory, the measure of the union of two sets is the sum of their measures, if and only if the sets are disjoint.

In the present experiment an attempt was made to produce two disjoint sets of cues, so that additivity of their measures would result. The theory used asserts that the rate of learning (θ) depends directly on the proportion of relevant cues, so that from additivity of cues one can predict additivity of learning rates. Furthermore, the theory can be used to estimate the learning rates, so that numerical predictions can be made. Provided that the sets of cues are in fact disjoint, these predictions constitute a test of the correctness of the theory.

The cues in the present problem arise from a pattern of letters (ABDEF). One problem is constructed where the letter A is relevant, a second problem where the letter E is relevant, and a third problem where both A and E are relevant and redundant. To make a given letter relevant, the correct response is made contingent on whether the letter appears capital or small. For example, if A is the relevant letter, *S* learns that when A appears capital (A), *X* is the correct response and when A appears small (a), *O* is the correct response. When a letter is irrelevant, the correct response is independent of whether that letter appears capital or small.

The model used to analyze the problems may be defined employing some of the elementary notions of mathematical set theory. When the letter A appears, on successive trials, in both capital and lower case, a variety of cues result. Some of these cues depend upon the contrast between the capital and lower-case form of the letter. These "contrast" cues are relevant when that letter is made relevant, as when A in the stimulus is followed by *X* and a by *O*. Such cues from the letter A are called *A*, and contrast cues from the letter E are called *E*. The letter in the stimulus may give rise to other cues which are, in the problems given in this experiment, always irrelevant. All irrelevant cues in the problem, from all sources, are assembled into one set, called *I*. If the letter A appears as a capital letter in all stimuli, then cues depending on the contrast of capital and lower-case form do not exist in the problem, though other cues arising from the letter A are presumably present. Thus, fixing a letter as always capital (or always lower-case) removes just the contrast set *A*, but does not affect the irrelevant elements from the letter which are in set *I*.

The set of cues is assumed to consist

[1] This study is based on a thesis submitted to the College of Science and Arts of Michigan State University in partial fulfillment of the requirements for the MA degree. The writer is indebted to Frank Restle for advice and assistance throughout the course of the investigation.

of a finite number of elements, each element associated with a positive number to denote its "weight"; the measure of such a set is the sum of all these numbers (Bush & Mosteller, 1951). The measure of A is written $m(A)$, and corresponds to the importance of this set of cues in controlling behavior. Similarly, the measures of E and I are written $m(E)$ and $m(I)$ respectively.

The union of two sets, written $A \cup E$, is the set containing everything in either A or E or both. The assumption made is that A and E are disjoint, having no common elements. In this case,

$$m(A \cup E) = m(A) + m(E)$$

In Problem 1, A is relevant, E is irrelevant, and *all other cues in the situation are irrelevant.* Then since θ is the proportion of relevant cues (Restle, 1955)

$$\theta_1 = \frac{m(A)}{m(A) + m(E) + m(I)} \quad [1]$$

In Problem 2, E is relevant and A is irrelevant,

$$\theta_2 = \frac{m(E)}{m(A) + m(E) + m(I)} \quad [2]$$

In Problem 3, both A and E are relevant and redundant, so that

$$\theta_3 = \frac{m(A) + m(E)}{m(A) + m(E) + m(I)} \quad [3]$$

Combining Equations 1 and 2 one sees that $\theta_1 + \theta_2 = \theta_3$. This is Hypothesis 1.

The set of E cues (relevant in Problem 2) depends on the fact that the letter E varies from capital to lower case. If we fix the letter by making it capital E on every trial, the set E has been removed.

In Problem 4, A is relevant, E is fixed, then

$$\theta_4 = \frac{m(A)}{m(A) + m(I)} \quad [4]$$

Similarly, in Problem 5, E is relevant, A is fixed, then

$$\theta_5 = \frac{m(E)}{m(E) + m(I)} \quad [5]$$

Solving Equations 4 and 5 for $m(A)$ and $m(E)$ and substituting these values into Equation 3, we find

$$\theta_3 = \frac{\theta_4(1 - \theta_5) + \theta_5(1 - \theta_4)}{(1 - \theta_4\theta_5)} \quad [6]$$

This is Hypothesis 2.

Both hypotheses have been used in the analysis of additivity of cues problems in several papers (Restle, 1955, 1957, 1958) but none of these applications was to problems specifically formulated to test the additivity hypothesis. Recently, Restle (1959) has tested direct additivity of learning rates (Hypothesis 1) using a pattern of consonant letters. This paper formulates these hypotheses using set theory and tests them on data well fitted for the purpose.

Method

Subjects.—The *S*s were 80 students in elementary psychology courses at Michigan State University who received credit for taking part in experiments. There were five experimental groups of 10 men and 6 women each.

Apparatus.—From the point of view of *S*, the apparatus consisted of a shield (9 in. × 18 in.) with two apertures (½ in. × 3 in.), one on the left and one on the right, and with a lever on the left. Stimulus patterns appeared in the left aperture, and responses (an *X* or an *O*) were recorded by *S*, in pencil, through the right aperture. The lever permitted *S*, after responding, to view the reinforcement (an *X* or an *O*) in the left window and then to obtain the next stimulus, the pace being controlled entirely by *S*.

Procedure and stimuli.—If more than one *S* was present, cardboard dividers were used to shield each *S* from view of the others' apparatus.

The same instructions were given to all groups. The *S*s were told that when they pushed the lever a group of letters (ABDEF) would appear in the window on the left. The same letters would appear every trial, but they would vary between upper and lower case. The correct answer, an *O* or an *X*, depended upon which letters were capital and

which were small. Responses were written only when the stimulus letters appeared.

All groups were given 128 training trials. The first trial was given as a practice trial to all *Ss* during the instruction session. This trial was not included in the recording of errors. Omitted responses on other trials were considered as one-half error.

Five problems were prepared, all involving the same letters, ABDEF, the letters varying between upper and lower case over trials, but retaining the same alphabetical order. For all problems, the letters B, D, and F were varied independently of one another and were uncorrelated with the correct response. Patterns of all capital or all small letters were omitted.[2] Problems 1 and 2 involved all of the remaining 30 letter patterns. Problems 4 and 5 included 15 different letter patterns, one letter always appearing capitalized. Problem 3 also involved 15 patterns, two of the letters being relevant so that half their combinations had to be eliminated.

Care was taken that the stimulus patterns appeared about equally often and in an irregular order. Consecutive runs of *X*s and *O*s were distributed as the expectation of a random 50:50 sequence.

Experimental groups.—Group A had Problem 1, in which capital A was always followed by an *O* and small a by an *X*. The letter E was irrelevant.

Group E had Problem 2, in which capital E was always followed by an *O* and small e by an *X*. The letter A was irrelevant.

Group A + E had Problem 3, in which both A and E were relevant and redundant. Small a and capital E (aE) were always followed by an *O* and capital A and small e (Ae) were always followed by an *X;* AE and ae did not occur.

Group A′ had Problem 4, which was the same as that of Group A except that the letter E always appeared capitalized.

Group E′ had Problem 5, which was the same as that of Group E except that the letter A always appeared capitalized.

Results and Discussion

Figure 1 gives the group performance curves in terms of mean proportion of correct responses per 16-trial block.[3] Group A + E learned fastest, suggesting the advantage of combined cues. Group A′ learned slightly faster than Group A and Group E′ learned slightly faster than Group E, suggesting that fixing an irrelevant dimension speeded learning of the other dimension. Groups A and A′ learned faster than Groups E and E′, indicating that the letter A at the left of the stimulus pattern was more salient than the letter E in fourth position. A simple analysis of variance was performed on the total error scores of the experimental groups and differences were found to be statistically significant ($F = 6.23$, 4/75 *df*, $P < .01$).

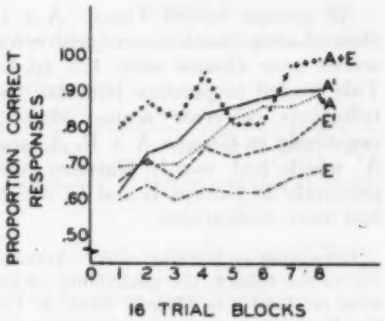

Fig. 1. Mean proportion of correct responses as a function of training trials for Groups A, E, A′, E′ and A + E.

[2] These extreme patterns were quite easily recognized and rapidly learned in previous studies using similar procedures. (Restle, Personal communication. 1959.)

[3] The initial running of Groups A and A′ gave an anomalous result in that Group A learned faster than Group A′. This difference was not significant at the .05 level, but the direction of the difference was unexpected. Both groups had the same relevant cue *A*, the only difference being that in Group A′ the irrelevant letter E was fixed. Group E′ learned faster than E, so the unexpected result was not repeated within the experiment. A recheck of experimental conditions failed to explain the difference. Groups A and A′ were rerun under the same experimental conditions in order to see whether or not this result would repeat itself. The second running of the groups did not show the same results and the data from the second two groups are used in the analysis.

All groups except Group A + E showed some "nonlearners" with error scores near chance over 128 trials. This tended to produce bimodal distributions of error scores, skewed negatively in Groups A + E, A, and A′ which had mostly learners, and positively in Groups E and E′ which had more nonlearners.

Estimation of learning rates.—According to the theory, the probability of an error on Trial n is (Restle, 1955, p. 13, Eq. 7)

$$1 - p_n = \frac{.5(1\theta -)^{n-1}}{\theta + (1 - \theta)^n}$$

Let X_n be a variable which has value 1 if S makes an error on Trial n, and 0 if S makes a correct response on Trial n. Then one estimate of θ is from total errors; the value of $\hat{\theta}$ is chosen which satisfies the equation[4]

$$\sum_{n=2}^{128} X_n = \sum_{n=2}^{128} (1 - p_n)$$

$$= \frac{.5 \log \hat{\theta} + (1 - \hat{\theta})^{128}}{(1 - \hat{\theta}) \log (1 - \hat{\theta})}$$

Another estimate is from weighted errors, being that value of $\hat{\theta}$ which satisfies

$$\sum_{n=2}^{128} (n - 1)X_n = \sum_{n=2}^{128} (n - 1)(1 - p_n)$$

$$= \sum_{n=2}^{128} \frac{.5(n - 1)(1 - \hat{\theta})^{n-1}}{\hat{\theta} + (1 - \hat{\theta})^n}$$

The term on the right of the equation has no closed solution and has been tabulated for these purposes. Both estimates were computed from the average values of $\sum X_n$ and $\sum (n - 1)X_n$ for each group. The results are shown in Table 1.

In Table 1, the estimates of θ based on weighted errors are consistently smaller than the estimates based on total errors.

[4] The notation $\hat{\theta}$ means the estimated value of the parameter θ.

TABLE 1

θ ESTIMATES FOR THE FIVE EXPERIMENTAL GROUPS

Group	Source of Estimate	
	Mean Errors	Mean Weighted Errors
A + E	.100	.070
A′	.060	.050
A	.055	.046
E′	.046	.039
E	.033	.029

Such a discrepancy suggests that the learning curves found did not have the shape stated by the theory. Since the estimates were based on group averages, the discrepancy might mean that individual differences were sufficient to distort the shape of the learning curve, or that individual *S*s did not distribute their errors as the theory predicts. Weighted error estimates place emphasis on errors made late in learning. From this, it appears that the groups made a disproportionate number of errors late in learning, compared with the theoretical expectations. This effect may be due to the presence of nonlearners.

To test whether individual *S*'s learning curves were of the shape predicted by

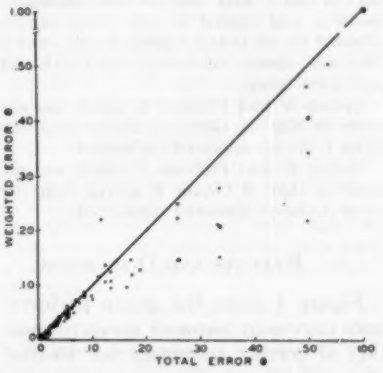

FIG. 2. Scatterplot of individual estimates of θ based on total error and weighted error scores. The large dots represent coincident observations.

the theory, individual $\hat{\theta}$ values were calculated for each S. If each S distributes his errors appropriately, then the two estimates of θ, from total and weighted errors, should be about equal for each S. An over-all sign test between the two sets of estimates showed that the individual weighted error estimates were consistently smaller than the individual total error estimates ($z = 3.91$, $P = .0001$).

The individual estimates of θ based on total error and weighted error scores are shown in Fig. 2. The discrepancies (indicated by deviations of the points from the diagonal line) are largest at high values of θ, which corresponds to rapid learning and very small error scores. The results are those to be expected if Ss make occasional late errors after learning is essentially complete. Fatigue, boredom, or occasional interruptions or noises, not accounted for in the theory, may have produced this effect. The numerical differences are small: in 80 comparisons, 70% of the discrepancies were less than .015.

Additivity of cues.—If the sets A and E of relevant cues are disjoint, then θ_1 and θ_2 should add directly to yield θ_3, since Problems 1, 2 and 3 all have the same cues. Similarly, θ_4 and θ_5 should combine by Equation 6 to yield θ_3.

To determine whether the estimates of θ in Table 1 are consistent with these experimental hypotheses, the method of prediction was used. Estimates of θ_1 and θ_2 were added to produce a prediction of θ_3. From this predicted θ_3, mean error and mean weighted error expectations were computed. These predictions are compared with the obtained values for Group A + E in Table 2. Because of the discrepancies between estimates, the procedure is repeated separately for mean errors and mean weighted errors.

Similarly, estimates of θ_4 and θ_5 are combined by Equation 6 to produce another prediction of θ_3. This predicted θ_3 is used to generate predicted error and weighted error scores which are compared with the obtained values in Table 3.

The t tests reported in Tables 2 and 3 are computed by taking the predicted values as fixed points. This test is overly stringent, since these predictions have an unknown variance resulting from variance of Groups A, E, A′ and E′.

TABLE 2

PREDICTIONS OF A + E FROM A AND E: HYPOTHESIS 1

Score	Pred.	Obs.	t	P
Mean errors	14.57	11.72	0.81	>.05
Mean weighted errors	428.00	495.34	0.35	>.05

In none of the four predictions of the performance of Group A + E (from Groups A and E with total errors, and from A and E with weighted errors, from A′ and E′ with total errors, and from A′ and E′ with weighted errors) were the predictions significantly in error. None of these predictions, taken as a fixed parameter, could be rejected at the .05 level.

As individual differences were present in terms of performance in the experimental groups, a further test of the predictions was carried out. The distribution of scores can be predicted for any assumed correlation between the scores on the two predictor problems. The assumption was made that there was a high positive correlation as to rank position in the different experimental groups. Acting upon this assumption, the individual θ values were ranked in each group. The matched rank θ values for Groups A and E were then added directly to yield a predicted distribution (Hypothesis 1). Similarly, the matched rank θ values for Groups A′ and E′ were combined by Equation 6 to yield a

TABLE 3

PREDICTION OF A + E FROM A′ AND E′: HYPOTHESIS 2

Score	Pred.	Obs.	t	P
Mean errors	11.72	11.72	0.00	>.05
Mean weighted errors	305.00	495.34	0.98	>.05

second predicted distribution (Hypothesis 2). These calculations were carried out using both total and weighted error estimates. Predicted distributions were compared with the obtained distribution of Group A + E by means of the Kolmorgorov-Smirnov cumulative distribution test for two observed distributions. A maximum difference of 8 is necessary for rejection of the hypothesis, whereas the differences ranged from 3 to 5, no difference approaching the .05 level. Furthermore, t tests between predicted and obtained distributions were all far short of significance (.05 level). Median-split chi square tests comparing each predicted distribution with the obtained distribution were performed, and all such tests failed to disclose a difference in medians (.05 level).

In this study, the difficulties encountered in finding comparable estimates of θ by the two methods suggested by the theory forced the making of predictions of additivity with separate θ values. Investigation of the two estimates on each S indicated that the discrepancy became larger as the rate of learning increased (see Fig. 2). The discrepancy in estimates was further investigated by Monte Carlo computations. The theoretical curve, $1 - p_n$, was computed on hypothetical data, exactly fitting the theory for $\theta = .25$ for 80 *Ss*. The parameter was then recovered by both estimation methods for each hypothetical S. It was found that the estimates were quite variable, with values centering in the region of .25. In the Monte Carlo procedure, the weighted error estimates gave higher estimates ($z = 3.15$, $P = .002$) but in the experimental data, the total error methods gave higher estimates. These results suggest that the theoretical learning curve does not fit the empirical data correctly. The exact source of the discrepancy is not known at this writing.

Summary

Four groups of 16 college students each learned problems involving patterns of five letters, in which only one of two key letters was relevant. For two of the groups the other letter was irrelevant; for the other two groups the other letter was fixed in appearance, reducing irrelevant cues. A fifth group of 16 *Ss* learned a problem with both letters relevant and redundant.

The two-letter-relevant problem was learned faster than any of the others, one of the cues was found to be stronger than the other and a small advantage due to reducing irrelevant cues was suggested by the findings.

Two hypotheses concerning additivity of cues were formulated and tested. The hypotheses differed with respect to details of experimental procedure and involved additivity of learning rates.

Learning rates (theoretically, the proportions of relevant cues) were estimated in problems where only one letter or the other is relevant and then used to predict performance on the problem in which both letters are relevant and redundant. Statistical tests failed to reveal any significant differences between prediction and observation in the mean, median, or over-all distribution of error scores.

These results were taken as supporting the assumption that the rate of learning depends on the proportion of relevant cues approximately as specified in Restle's theory.

References

Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.

Restle, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

Restle, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64**, 217–228.

Restle, F. Toward a quantitative description of learning-set data. *Psychol. Rev.*, 1958, **65**, 77–91.

Restle, F. Additivity of cues and transfer in discrimination of consonant clusters. *J. exp. Psychol.*, 1959, **57**, 9–14.

(Received August 1, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 89–96

EFFECT OF INTERMITTENT REINFORCEMENT ON ACQUISITION, EXTINCTION, AND SPONTANEOUS RECOVERY OF THE CONDITIONED EYELID RESPONSE [1]

THOMAS F. HARTMAN AND DAVID A. GRANT

University of Wisconsin

The purpose of the present experiment was to investigate the effect of various degrees of intermittent reinforcement upon the acquisition, extinction, and spontaneous recovery of the conditioned eyelid response when a constant number of reinforcements was given. The research investigates two aspects of conditioning that are of special interest. First, although it was evident from the study of Grant and Schipper (1952) that increasing percentages of reinforcement result in increasingly higher asymptotes during acquisition of the conditioned eyelid response, the function was not investigated for the situation in which the number of reinforcements was equalized over the whole acquisition process for each percentage of reinforcement. Secondly, there are very few spontaneous recovery data on human *S*s in the literature (Ellson, 1939; Grant, Hunter & Patel, 1958; Howat & Grant, 1958; Porter, 1938, 1939) and the problem of spontaneous recovery after intermittent reinforcement has been totally ignored. The two most general attempts to establish a theoretical framework for these phenomena, the Hull behavior theory approach (Hull, 1951) and the Estes probability model approach (Estes, 1950, 1955; Estes & Straughan, 1954), do not deal explicitly with all of the phenomena involved, but the availability of empirical data may serve to indicate the ways in which these formulations may be modified.

METHOD

Apparatus.—The apparatus was modified from that used in earlier research (Grant, Schipper, & Ross, 1952). The *S* sat in a straight-backed chair facing a rectangular enclosure placed on top of a table. The interior of the box was painted white. About 80 cm. in front of *S*'s eyes and slightly below his line of level vision was a white rectangular box with a 10-cm. circular milk glass window in which the conditioned stimulus (CS) was presented. The background illumination in the enclosure varied in the neighborhood of 5–10 mL., and the CS consisted of illumination of the window to 42 mL. for 750 msec. The *S* wore a 5-cm. × 15-cm. sheet of 28-gauge rubber-padded aluminum on his face lateral to his left eye. This plate was supported by an elastic band around his head and held the stimulus jet and a microtorque potentiometer[2] which served to pick up *S*'s eyelid movements. A hypodermic tube was fitted to the shaft of the potentiometer and a length of piano wire was passed through the tubing so that it slid smoothly. A small piece of thin aluminum was attached to the end of the piano wire, and this in turn was taped to *S*'s eyelid by means of Scotch Tape so that any movement of *S*'s eyelid would be transmitted to the shaft of the potentiometer. Timing and recording equipment was electronic. The signal from the potentiometer was amplified by a dc amplifier and differentiated once so that the first time derivative and the actual eyelid movements could be recorded by means of an ink writer in *E*'s

[1] This research was supported in part by the National Science Foundation and in part by the Research Committee of the Graduate School of the University of Wisconsin with funds from the Wisconsin Alumni Research Foundation.

[2] A recording device originally suggested by K. W. Spence.

room. The CS and UCS were also indicated on the record.

On reinforced trials, the UCS was a corneal air puff of 200-msec. duration administered 500 msec. after the onset of the light. The puff involved a pressure of 150 mm. of mercury at the source and was sufficient at the eye to produce a consistent reflex. It was delivered through a plastic tube of 3 mm. aperture diameter approximately 9 mm. from the lateral portion of the cornea of *S*'s left eye.

Experimental design.—Each *S* was given 40 reinforced acquisition trials immediately followed by 20 extinction trials during the first experimental session which lasted about 40 min. Approximately 24 hr. later, a second session of 20 unreinforced trials was given to each *S* to test for spontaneous recovery. This session lasted approximately 10 min.

Four groups of 20 *S*s each were used in the present experiment. The 25% group received 160 acquisition trials with a mean intertrial interval of 12 sec.; the 50% group received 80 trials with a mean intertrial interval of 24 sec.; the 75% group received 54 trials with a mean intertrial interval of 36 sec.; and the 100% group received 40 trials with a mean intertrial interval of 48 sec. The average interval between reinforced trials was thus essentially the same in the four groups. Four independent random sequences were used with each of the intermittent groups with the restrictions that: the first and last acquisition trials were reinforced; one half of the reinforced trials came at the first half and one half at the last half of acquisition; and there were never more than six consecutive unreinforced trials in a series. Intertrial intervals and reinforcements were independently randomized and were programmed by means of a Western Union tape transmitter. In the extinction and spontaneous-recovery test series, the mean intertrial interval was approximately 30 sec. and contained equal numbers of the intertrial intervals present in each acquisition so that no group would receive intervals in extinction more like those in acquisition than any other group received.

Subjects.—The *S*s were 32 men and 48 women students who volunteered from classes in elementary psychology at the University of Wisconsin. They were assigned randomly within sex to the four experimental groups. The *S*s were instructed to maintain a passive attitude toward the experiment and not to influence in any way their natural reactions to the stimuli.

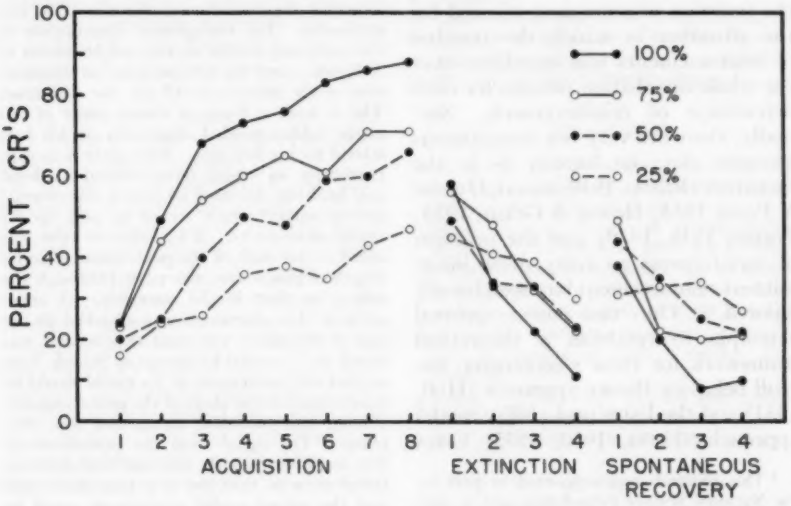

FIG. 1. Percentages of CRs plotted for successive eighths of acquisition and successive quarters of extinction and 24-hr. spontaneous recovery test with percentage of reinforcement during acquisition as the parameter.

Results

Owing to the general stimulus conditions of the experiment there were few signs of voluntary responding (Spence & Ross, 1959) among the *Ss*, and no *Ss* were rejected for this reason. All eyelid responses with latencies between 200 msec. and 550 msec. were defined as CRs. Figure 1 presents the acquisition, extinction, and spontaneous recovery findings of the experiment. The percentages of CRs are plotted against successive eighths of the acquisition series and successive quarters of the extinction and spontaneous recovery series. For the 100% group this would mean every five-trial block throughout the experiment, but the size of the acquisition block would vary according to the percentage of reinforcement during acquisition. The parameter is percentage of reinforced trials during the training period.

Inspection of Fig. 1 reveals immediately that even when the numbers of reinforced trials are equalized, the percentage of intermittence of reinforcement has essentially the same effect during acquisition, as was reported by Grant and Schipper (1952), when the numbers of reinforcement trials were not equal. The terminal levels during acquisition are very similar to those reported by Grant and Schipper except for the 75% group, which is considerably lower in the present experiment than in the earlier one. An analysis of variance based on the percentage frequency of eyelid CRs during the last 20 trials of acquisition yielded a highly significant F of 6.889 ($df = 3/76$), with an error MS of 516.399 for arcsin transformed percentages.

During extinction the trends are very orderly and show more rapid extinction with the higher percentages of reinforcement. In fact, at the end of the extinction session the tendency to give CRs is reversed with respect to the various percentages of reinforcement. Here again, the data are quite similar to those reported by Grant and Schipper, although in the present experiment the 25% reinforcement group gives much more resistance to extinction than was reported in the earlier experiment. This discrepancy presumably was due to the smaller actual number of reinforcements given in the 25% group of the earlier experiment. The arcsin transforms of the percentage frequency data on successive five-trial blocks during extinction were analyzed and compared for the four groups by an orthogonal polynomial trend test (Grant, 1956). A summary of this analysis is presented in Column A of Table 1. The significant over-all linear component simply indicates that the over-all average scores decrease from block to block and could be described essentially by a linear equation. Of considerably more interest are the significant between-groups trend and linear component. Here the lower percentage of reinforcement results in significantly less negative slope than the higher percentage of reinforcement; these *Ss* extinguish significantly more slowly, as indicated in Fig. 1. The significant individual differences in average performance shown by the significant F between individual means is a common finding.

The spontaneous recovery test data graphed in the right-hand of Fig. 1 are of the same basic form as the initial extinction curves and the first impression is that there is tremendous 24-hr. recovery for all groups. Actually, the ordering on the first block of spontaneous recovery testing

TABLE 1

TREND ANALYSES OF TRANSFORMED FREQUENCY OVER FIVE-TRIALS BLOCKS DURING (A) EXTINCTION AND (B) SPONTANEOUS RECOVERY

Source of Variation	*df*	A. Extinction *Fs*	B. Spontaneous Recovery *Fs*
Over-all trend	(3)	32.066***	21.013***
Linear	1	58.625***	35.772***
Quadratic	1	0.184	11.682***
Cubic	1	0.965	0.907
Between reinforcement percentages	3	0.290	0.860
Between reinforcement trends	(9)	2.136*	3.462**
Linear	3	2.853*	3.411*
Quadratic	3	1.550	5.507**
Cubic	3	1.006	1.227
Between-*S* means	76	10.702***	9.493***
Between-*S* trends	(228)	(314.55)	(261.69)
(Linear error *MS*)	76	(511.99)	(387.10)
(Quadratic error *MS*)	76	(221.28)	(212.36)
(Cubic error *MS*)	76	(210.37)	(185.60)

* Significant at 5% level.
** Significant at 1% level.
*** Significant at 0.1% level.

is not quite the same as that in the extinction due to the depressed tendency of responding in the 100% reinforcement group. The extended trend analysis on the transformed spontaneous recovery test data in Column B of Table 1 shows significant results similar to those found in extinction, except that the quadratic components of the over-all trend and the quadratic components of the between-group trends are significant. The significant between-group quadratics presumably arise because of the greater curvilinearity of the spontaneous recovery extinction test in the 100% and 75% groups, as compared with the 50% and 25% curves.

To determine if the forms of the second extinction curves on the spontaneous recovery test were the same as for the initial extinction, an analysis of variance was carried out on the orthogonal trend components of the transformed scores summarized in Table 1. These results are given in Table 2. The extinction-spontaneous recovery *MS* is a conventional between-treatments *MS*, and the error *MS* was based on *Ss* × Treatments interaction. The error term for each *F* test is the extinction-spontaneous recovery interaction with *Ss*. Of the effects tested, only the difference between linear components of extinction and spontaneous recovery tests in the 25% group and the difference between quadratic components of the 75% group reached significance. That is to say, the second extinction curve of the 25% group was flatter than the

TABLE 2

TESTS OF INITIAL EXTINCTION VS. SPONTANEOUS RECOVERY TREND COMPONENTS

Group	Source of Variation	*df*	Linear Component *F*	Quadratic Component *F*
25%	Ext. vs. Spont. rec.	1	5.193*	1.220
	Error (*MS*)	19	(152.13)	(183.49)
50%	Ext. vs. Spont. rec.	1	0.057	1.239
	Error (*MS*)	19	(161.25)	(200.16)
75%	Ext. vs. Spont. rec.	1	1.086	8.656**
	Error (*MS*)	19	(285.49)	(229.67)
100%	Ext. vs. Spont. rec.	1	3.990	0.759
	Error (*MS*)	19	(286.90)	(177.80)

* Significant at 5% level.
** Significant at 1% level.

first, and the second extinction of the 75% group was more curved than the first extinction. The changes in shape of the other extinction curves were not significant.

The amount of spontaneous recovery for the various groups is somewhat difficult to ascertain, owing to the problem of selecting a suitable index. In general, a simple index is desirable, and when the percentage of response on the first block of trials during recovery is compared with the percentage of response on the first block of initial extinction, it is found that there is 69% recovery in the 25% group, 90% recovery in the 50% group, 88% recovery in the 75% group, and 66% recovery in the 100% group. Thus the 50% and 75% groups show greater spontaneous recovery than was found in the 25% and 100% groups.

Discussion

Acquisition.—In general, the results of Grant and Schipper (1952) and Ross (1959) are confirmed in that higher percentages of reinforcement result in higher levels of conditioning, and this finding is generalized to the situation where the number of reinforcements is equal for all groups. A decremental effect of the nonreinforced trials (Ross, 1959; Ross & Spence, 1960) is indicated. It will be remarked that the lower percentages of reinforcement involved higher degrees of massing than was the case for the 100% group. In general, massing with 100% reinforcement will depress acquisition (e.g., Spence & Norris, 1950); with 50% intermittent reinforcement, however, Grant, Schipper, and Ross (1952) found no significant effect of distribution, and it is conjectured that massing is more important with 100% reinforcement than with intermittent reinforcement, although this matter merits further investigation.

Since the acquisition curves did not reach definite asymptotes at the end of the acquisition training, the magnitudes of the discrepancies between the data and the percentage of reinforcement values (π), as predicted by the Estes-Straughan model (1954), are not determinate. The 25% and 50% groups appeared to be headed for asymptotes higher than the 25% and 50% π values. An indication of a departure from the elegant Estes-Straughan model seems to have been found, although the appropriateness of this model for eyelid conditioning may be questioned, since the eyelid CR is probably not as simple as the conditions envisaged by the model.

Extinction.—The extinction results are in line with the classical results in the literature (Grant & Schipper, 1952; Humphreys, 1939). An interpretation of the differential extinction found with varying percentages of reinforcement has not been extensively dealt with by either Estes and Straughan (1954) or by Hull (1951). In the Estes-Straughan formulation, it would appear that once a given level of habit strength is reached, extinction should be a function only of this level and the distribution of trials in the specific extinction schedule (Estes, 1955). In this experiment all groups had the same extinction series. The effect of distribution of acquisition trials upon initial extinction was not considered to be satisfactorily relevant to test the Estes theory. A family of extinction curves beginning at the terminal acquisition level and approaching a common asymptote at similar rates would be predicted by the Estes model. Such was not found, but rather a progressively greater resistance to extinction with lower percentages of reinforcement.

The simplest change in the Estes model that would help account for these results would be an assumption that intermittence tends to increase the interchange between elements in the available subpopulation, S, and the unavailable subset of stimulus elements, S′, in the superpopulation, S*, during both acquisition and extinction of the response. Such an addition to the theoretical formulation is plausible during acquisition, because the complexity of the

intermittence schedule is disruptive to *S* and might tend to bring new elements into his available stimulus set. The increased exchange in acquisition would allow a large number of elements in the superpopulation to be conditioned even though the percentage conditioned in the available subset might never be high. With less interchange in the continuously reinforced procedure the percentage conditioned in the available subset might be high without ever conditioning a large number of elements in the superpopulation. The extinction process after intermittent reinforcement would be prolonged since a large number of elements were conditioned, provided that it is also assumed that a relatively free interchange takes place between subsets during the extinction process. This causes difficulties later in connection with spontaneous recovery. Extinction after continuous reinforcement ought to be quite rapid as the conditioned elements from the unavailable subset would not be diffusing into the available subset at as high a rate as that found in the other procedures.

Spontaneous recovery.—The spontaneous recovery data are new to the literature and are noteworthy in that the degree of spontaneous recovery is high, as might be expected from continuous reinforcement data reported by Porter (1938, 1939), Howat and Grant (1958), and Grant, Hunter, and Patel (1958). The fact that even higher spontaneous recovery is found following 50% and 75% reinforcement is especially striking although statistical comparisons between groups were not feasible on this finding. Certainly the amount of spontaneous recovery was high for all procedures.

The high spontaneous recovery produces problems for the Hullian theory unless it is assumed that the initial extinction is caused primarily by I_R (Hull, 1951). The spacing of initial extinction trials, however, would seem to be adequate to keep I_R at a low level (Spence & Norris, 1950). The problem of reduced drive, *D*, (Spence, 1956, 1958) owing to the suspension of the UCS during extinction is a more complicated matter. An explanation of high levels of spontaneous recovery in terms of *D* evidently requires that the part of *D* that is based on conditioned emotional responses must recover over a 24-hr. period in which no additional UCSs are presented. This, unfortunately, explains spontaneous recovery of eyelid CRs in terms of the spontaneous recovery of other CRs, which may be stretching the *D* theory too far.

The spontaneous recovery results also pose special problems for the Estes spontaneous recovery and regression model (Estes, 1955). If the simple Estes model were valid for the present situation one would predict spontaneous recovery curves similar in form to the initial extinction curves but with lower initial response frequencies. The higher level of responding at the end of initial extinction by the groups receiving the lower percentages of reinforcement immediately complicates the predictions of spontaneous recovery from the Estes model as indicated in the previous section. It seems evident, however, that the crossing over of the curves observed in the spontaneous recovery test and the differences in trend components of extinction and spontaneous recovery tests would hardly be expected. If the tendency of the groups with the lower percentages of reinforcement to respond with a higher frequency of CRs during the beginning of the spontaneous recovery test series is a valid observation, further difficulties confront the model. Aside from this, however, if it be granted that the groups with low percentages of reinforcement during acquisition showed high spontaneous recovery, it is again indicated that greater diffusion of stimulus elements between the two subsets of the superpopulation, available and unavailable, would be required with increasing intermittence of reinforcement. Furthermore, because of the high degree of spontaneous recovery, it is necessary that the unavailable subset be large relative to the available subset. With relatively free interchange in the 25% group the available stimulus subset would never differ much from

the rest of the superpopulation during initial extinction, so that the spontaneous recovery curve would not show great initial recovery and only a small amount of negative slope or extinction. The continuously reinforced group, with the least interchange of elements, would present a markedly different available subpopulation after 24-hr. than was present at the end of initial extinction, since the extinction effects were confined to a restricted, well-defined group of stimulus elements during initial extinction, and the intervening 24-hr. should be sufficient, even with restricted exchange, to obtain a relatively new available subset. Spontaneous recovery, however, should not be great in this group, since the restricted interchange in acquisition would not allow a high percentage of elements in the superpopulation to be initially conditioned. The 50% and 75% groups with intermediate degrees of interchange of elements between the available and unavailable subsets may turn out to be optimal for spontaneous recovery.

Such a change in the Estes model, though simple and plausible from the psychological standpoint, poses grave mathematical problems in obtaining detailed predictions, because intermittence would have a progressively greater effect during acquisition. This results in nonlinear difference equations.

Summary

The effects of 25%, 50%, 75%, and 100% random reinforcement upon the acquisition, extinction, and spontaneous recovery of the conditioned eyelid response were studied in 80 human *Ss*. Each *S* received 40 reinforced trials followed by a 20-trial extinction series on Day 1. The duration of the acquisition series and the number of reinforced trials was the same for all groups. Approximately 24 hr. later all *Ss* received the 20-trial spontaneous recovery test. When both reinforced and unreinforced trials were given, the order was random with specified restrictions. The *Ss* were light-adapted and were tested with neutral instructions. The principal findings were as follows:

1. The percentage frequency of CRs during acquisition was a function of the percentage of reinforced trials as has been reported in earlier work where the number of reinforced trials was not equalized.

2. During extinction the groups with the lower percentages of reinforcement showed significantly more resistance to extinction than those with higher percentages of reinforcement.

3. The form of the spontaneous recovery function for each group was not always similar to the extinction function for that group, although during the spontaneous recovery test the groups with the lower percentages of reinforcement again showed the greatest resistance to extinction.

4. There was some indication that the degree of spontaneous recovery was greatest in the 50% and 75% reinforcement groups.

The results were discussed in relation to Estes' stimulus variability learning model and Hull's behavior theory. Both theoretical approaches have difficulty in accounting for extinction following intermittent reinforcement and the high degree of spontaneous recovery that was obtained. It was suggested that greater compatibility with the Estes model could be obtained by hypothesizing that intermittent reinforcement tends to promote increased interchange between the elements in the available subpopulation and the remainder of the superpopulation of stimulus elements during both acquisition and extinction.

References

Ellson, D. G. Spontaneous recovery of the galvanic skin response as a function of the recovery interval. *J. exp. Psychol.*, 1939, **25**, 586–600.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves *Psychol. Bull.*, 1956, **53**, 141–154.

Grant, D. A., Hunter, H. G., & Patel, A. S. Spontaneous recovery of the conditioned eyelid response. *J. gen. Psychol.*, 1958, **59**, 135–141.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the per-

centage of fixed ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

GRANT, D. A., SCHIPPER, L. M., & ROSS, B. M. The effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

HOWAT, M. G., & GRANT, D. A. Influence of intertrial interval during extinction on spontaneous recovery of conditioned eyelid responses. *J. exp. Psychol.*, 1958, **56**, 11–15.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

PORTER, J. M., JR. The modification of conditioned eyelid responses by successive series of nonreinforced elicitations. *J. gen. Psychol.*, 1938, **19**, 307–323.

PORTER, J. M., JR. Experimental extinction as a function of the interval between successive non-reinforced elicitations. *J. gen. Psychol.*, 1939, **20**, 109–134.

ROSS, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

ROSS, L. E., & SPENCE, K. W. Eyelid conditioning performance under partial reinforcement conditions as a function of UCS intensity. *J. exp. Psychol.*, 1960, **60**, 379–382.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., & NORRIS, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

SPENCE, K. W., & ROSS, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received August 7, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 97–102

RESPONSE STRENGTHENING BY INFORMATION AND EFFECT IN HUMAN LEARNING

W. A. HILLIX AND MELVIN H. MARX

University of Missouri

The nature of the action of response consequences is one of psychology's persistent problems. Several investigators (e.g., Eisenson, 1935; Postman & Adams, 1954; Thorndike, 1933, 1935) have held that the action of reward is essentially automatic and noncognitive. Another group (e.g., Bitterman, 1956; Tolman, Hall, & Bretnall, 1932) has held that the strengthening observed is due to the information given by the aftereffect. Experimental evidence is available which appears to support each point of view. For example, Muenzinger (1934) showed that information can strengthen connections for human *S*s independently of any experimentally manipulated satisfying aftereffects. On the other hand, Porter (1957) recently found evidence apparently confirming an automatic strengthening view.

It is, of course, quite possible that both information and effect are sufficient to strengthen responses. In any case, it is difficult to present effect without also introducing information. The problem we are concerned with, therefore, is not information versus effect, but rather the role of effect when added to that of information alone.

The original purpose of the present study was to test the proposition that when responses are followed by reward they are more strongly established, as measured by a transfer test, than when they are followed by an equivalent amount of information provided in the absence of the presumed "effect" function. It was assumed that the two training procedures would result in essentially similar acquisition performance, and that if there were any difference it would be in favor of the experimental condition.

METHOD[1]

Subjects

The *S*s were 60 students from an undergraduate psychology class at the University of Missouri. Twenty served in each of the two groups in Exp. 1, 10 in each group in Exp. 2.

Apparatus

The apparatus was a Masonite box in which lights and switches were mounted. The front panel was 26 in. high and 38 in. long. Six rows and six columns of red jewel lights were mounted on it in a rectangular array. A toggle switch was beside each light. The lights and associated switches were numbered from 1 through 36 in normal English reading order. The back panel of the apparatus, which was seen and used only by *E*, consisted of switches connected in series to the front switches. Only if the appropriate switch on the back was closed would the front switch turn on the light next to it. Thus, *E* could choose any arbitrary pattern of lights which would come on when the switches were thrown on the front panel.

A larger green jewel light was mounted separately in a small Masonite box which sat on the top of the larger one. Rotary switches on *E*'s panel were connected with the front switches so that one of the front switches could be selected to turn on the green jewel light.

Procedure: Experiment 1

Acquisition trials.—All *S*s were run individually in sessions lasting about 1.5 hr.

[1] Further details and illustrative diagrams, as well as individual data, are available in the unpublished MA thesis (Hillix, 1954) from which the present report was prepared.

They were given diagrams with six rows and six columns of dots corresponding to the lights on the apparatus. Each of the five diagrams indicated connections between the dots, with a single dot in the left-hand column connected by lines to several dots in the second column; each dot in the second column was connected to more than one dot in the next column, and so on to the last column. The total number of dots so connected in any single diagram varied from 14 to 19.

Each experimental *S* received information plus reward (IR). He was told that *E* would turn on a light in the left-hand column, and that he should then turn to the diagram which began with that particular light. His task was then to guess which light was correct in each succeeding column and to tell *E* to check the guess by flipping the switch by the light. Correct lights came on, but incorrect ones did not. Switches were returned immediately to the off position after each check, regardless of outcome.

The experimental *S* continued to guess from among the possibilities indicated on the diagram until the correct light was found in a given column before proceeding to the next column. Choices in any given column were limited to those lights connected with the correct light in the preceding column of the diagram. There were from one to three lights so connected to each correct light in Exp. 1. Once a correct light had been discovered, *S* was instructed to remember it so that he could find all the correct lights immediately on later trials.

Control *Ss* received only information (I). They were instructed to learn the circuits by watching *E* throw switches and remembering the lights that came on. They were to call out the number of each switch *E* tried, whether right or wrong, so that *E* would be sure that the switch had been seen correctly. This procedure insured that *Ss* in both groups would make the same overt responses. For each *S* in the I group, *E* tried all the switches, both right and wrong, which the preceding *S* in the IR group had instructed him to try, in the same order and at the same rate as regulated by the buzzer. Information and overt responses were thus equated between groups. The critical difference between the I and IR groups was that IR *Ss* selected the switches to be tried in each column, while I *Ss* only observed the switches selected by *E* and called out their numbers.

Practicing each of the five training circuits one time was considered one training trial. The *S* was told to learn only the lights that came on, and to learn them well enough so that he could write the numbers of all the five lights that were correct within a given circuit, once *E* had turned on the light in the left-hand column which indicated the circuit number. For the acquisition test, *E* gave *S* a sheet of paper with blanks for each of the columns. Then *E* turned on the first light of a circuit and asked *S* to write the numbers of the rest of the lights that had been correct during training.

Circuits were presented in different random orders for each training and test trial. Rate of responding was fixed for both training and test by instructing *S* to respond immediately after the sounding of a buzzer which was set to operate every 5 sec. All *Ss* were trained and tested on a sample circuit to make sure they understood the instructions and became acclimated to the response rate. Acquisition tests were given following the Training Trials 3, 5, and 6.

The *E* motivated *Ss* in both groups by telling them in advance that they would be tested. Before the experiment began *Ss* took a test over a sample circuit. Also, *E* told each *S* that he would compare him directly with an *S* in another group, and that *E* would like to see him do better than the *S* in the other group. Work by Sims (1928) indicated that this second method should be effective, and it seemed to be.

Transfer trials.—These trials were identical for both groups. The *Ss* were told that the interconnections on their diagrams represented networks consisting of the previously learned circuit with additional parts; and that the networks were designed so that "shorts" could be introduced at some switch positions. The *S*'s task was to find the positions of these shorts. A short was indicated by the left-hand column light and one additional light from the circuits. Each left-hand column light was paired with each of two lights from the previous circuit. For one of these pairings the short was in the previously learned circuit; for the other, it was not. Thus, on the first trial, *S* turned to the diagram indicated by the light in the left-hand column and checked by throwing switches. When the correct switch was thrown, the associated light came on, and the green jewel light on the top of the board also came on. On succeeding trials, *S* was supposed to remember the position of the short within the appropriate network associated with the paired indicator light. Four trials were given on each of the 10 shorts (two within each of the five networks).

The instructions were designed so that *Ss* should have had no more reason to repeat

previously learned lights than to respond at new positions; *Ss* were explicitly told that the short was as likely to occur in the previously unlearned part as in the previously learned part. Motivation to repeat responses already learned for a given circuit should have been no greater than that to make new responses, and in any case should not have been different for I and IR *Ss*.

Procedure: Experiment 2

Certain unexpected acquisition results which occurred in Exp. 1 could have been a function of differential amount of time spent in following the diagrams, since IR *Ss* had to check them more closely than I *Ss*. As a matter of fact, two of the latter ignored the diagrams completely and spent their time in rehearsal; other I *Ss* referred to them only in a cursory manner. To eliminate this possible artifact, a procedural change was initiated in Exp. 2, which was designed to verify the earlier acquisition results. Diagrams were not used in the acquisition phase. Instead, *Ss* were told that circuits were constructed so that the correct light in any column was either directly to the right, to the right and just above, or to the right and just below, the correct light in the previous column. This gave *Ss* three choices in nearly all cases in Exp. 2, compared to the two choices that were most common in Exp. 1. The same lights were correct in each circuit that had previously been correct, and the rest of the procedure was identical to that of Exp. 1.

Results

Acquisition Trials

In Exp. 1, the mean number of correct responses per test was 17.62 for the IR group and 20.38 for the I group, out of a possible 25 correct responses. The variances were heterogeneous; after the Cochran-Cox correction (Edwards, 1950, p. 168) was applied, the difference between groups remained significant ($F = 7.72$, $df = 1/19$, $P < .05$).

In Exp. 2, the mean number correct was 17.37 for the IR group, and 20.83 for the I group. The variances were not heterogeneous, and the group difference was again significant ($F = 6.17$, $df = 1/18$, $P < .05$).

The Group by Trials interaction was not significant in either experiment.

Transfer Trials

If a given *S* had a greater tendency to respond in the learned part of the circuit when searching for "shorts," he should find them more quickly when they occurred in the previously learned part. It was predicted that IR *Ss* would have greater strengthening of S-R connections, and thus do better on shorts within the learned part of the circuit. Conversely, the I *Ss* would do relatively better on those in the unlearned part, since they would tend to distribute their guesses more evenly between learned and unlearned parts. The over-all effect should be reflected in an interaction between groups and learned-unlearned classification of the short location.

An analysis of variance was run on number of responses to find the short, with learned or unlearned position of the critical switch as one of the variables. The Group × Learned-unlearned position interaction, for all four trials on a given problem combined, closely approached significance at the .05 level in Exp. 1 ($F = 3.89$, $df = 1/38$, $.05 < P < .06$). The interaction was in the predicted direction; the IR group required 3.79 checks to find shorts in the learned part and 4.19 checks in the unlearned part, while the direction was reversed for the I group (4.52 vs. 4.31).

The group difference in number of checks was not significant when tested by analysis of variance; however, a simple analysis of covariance with total number of correct responses during acquisition as the control variable showed that the adjusted means of 144.94 for the IR group and 191.09 for the I group were significantly different ($F = 8.16$, $df = 1/37$, $P < .01$).

In Exp. 2, there was no significant group difference or Group × Learned-unlearned interaction.

A second measure of tendency to check in the learned part of the circuit was how many of the first five responses made in looking for each short were in the learned part of the circuit. There was no group difference within either experiment on this measure, but a difference did appear *between* experiments; in Exp. 1, a mean of 1.77 out of the first five responses were in the learned part; in Exp. 2, only 1.54 were in the learned part. The difference was significant ($t = 6.16$, $df = 58$, $P < .01$).

Discussion

Acquisition.—The superiority in acquisition of Group I was not predicted and was contrary to the general rationale of the study. However, confirmation of the significant difference in Exp. 2 indicated that the result could be accepted with some confidence. Even those who have argued that the important thing in acquisition is the informative value of the aftereffect would seem to have no basis for predicting the differential learning that occurred in the present studies, for a carefully equated amount of information was given to all *S*s.

It is commonly assumed (e.g., Hovland, 1951; Hovland, Lumsdaine, & Sheffield, 1949) that more active conditions produce better learning. However, the IR *S*s in the present study were more active with the diagrams in Exp. 1 and were certainly more active in the "decision making" sense. Their inferiority in learning indicates that the activity variable may be less important than is generally supposed. In fact, there is surprisingly little controlled experimental evidence for the generally beneficial effect of activity in human associative learning. The present results indicate a need for investigation of which kinds of materials are more effectively learned in a relatively passive manner through the presentation of information alone. It may be that wherever the motor components of the learning process are relatively simple or already well learned, there is little need for activity.

Although no perfectly clear explanation of the acquisition difference is readily available, certain interesting speculations can be made, based upon qualitative observations of *S*s' behavior by *E* as well as a simple logical analysis of the situation. Two major classes of conditions may be suggested, both relating to the greater activity, in the decision making sense, exercised by the IR *S*. In the first place, errors made by *S* himself may tend to be more resistant to elimination. A related bit of evidence here is the earlier finding (Farber, 1940; Marx, 1941; Stone, 1950; Tuckman, 1933) that the probability of repetition of an error, in multiple-choice human learning, is an increasing function of the number of times it has already been successively repeated. If errors in the present study were thus generally stronger for IR *S*s, they would have tended to produce greater interference with fixation of correct responses. In the second place, the need to decide which response to make very likely filled the interresponse interval of 5 sec. more completely for the IR than for the I *S*. Although this would be especially true in Exp. 1, where the diagrams were consulted, it would also be true to some extent in Exp. 2, where the decision making response was still necessary. In addition to providing some additional time for rehearsal for the I *S*, even in Exp. 2, this difference seemed to place the IR *S* under something of a "strain" as indicated by certain spontaneous responses made by such *S*s; any such mild emotional arousal might have deleterious effects on acquisition. Although the relevance of these variables to the present experiment can not be ascertained, they offer possibilities for experimental manipulation in future research.

The above considerations raise a

number of questions concerning the role of activity in learning, especially as it relates to the old problem of guidance. Guidance studies (e.g., Carr, 1921; Haslerud, 1953; Waters & Ellis, 1931) have typically not equated some of the variables that would be expected to influence performance; errors, time, or trials may vary, and the experiments thus do not constitute anything like a clean test of the effects of guidance per se. The present experiments can be considered as providing such a test, at least insofar as the concomitant variables mentioned above were controlled. In the typical guidance experiment, errors are simply prevented, and one cannot evaluate the relative effects of the guidance and the differential errors or concomitant time variations. Further research is needed to study the effects of varying the number and manner of introduction of errors.

Transfer.—There is some evidence in support of the original prediction, for the Group × Learned-unlearned interaction in Exp. 1 closely approached the .05 level of significance. The fact that this result was not duplicated in Exp. 2 weakens our confidence in the hypothesis. However, our evaluation of such negative evidence is tempered by at least two *ad hoc* considerations which suggest that Exp. 2 offers a less adequate test of the hypothesis. First, it had only half the number of *Ss* used in Exp. 1 (being primarily designed to verify the unexpected acquisition results). Second, the absence in Exp. 2 of the diagrams during acquisition may well have seriously reduced the transferability of the fixated responses by removing an important part of the stimulus complex that was present in Exp. 1. The importance of the diagrams is suggested by the superiority of IR *Ss* to I *Ss* in the transfer test in Exp. 1; presumably greater familiarity with the diagrams was obtained during acquisition for these *Ss*. Also, *Ss* in Exp. 1 responded significantly more often in the previously learned parts of the circuits than *Ss* in Exp. 2, again suggesting the importance of the diagrams as mediating stimuli.

Summary

Sixty college students took part in two experiments designed to examine certain aspects of the roles of information and reward in multiple-choice learning. In each experiment, one group (IR) learned patterns of lights with reward ("right" signal) following responses which they determined; a second group (I) learned the lights by receiving equivalent information as to correctness following the same responses determined by *E* and merely observed by *S*.

The experiments included both acquisition and response-transfer tests. The most interesting result was the unexpected finding that *Ss* receiving information alone learned significantly better on acquisition trials in both experiments than *Ss* receiving information plus reward. The prediction that IR *Ss* would do relatively better than I *Ss* on response-transfer trials where the problem involved the previously learned part of the circuit was supported at the .06 level of significance in Exp. 1, but was not supported in Exp. 2.

The inferiority of the IR *Ss* in acquisition was tentatively suggested to be a function of several possible factors, such as greater distraction by errors. The results indicate a need for a re-examination of the role of "activity" in human learning, especially as it relates to guidance, since the IR *Ss* were more "active" in a decision making sense. The original hypothesis that reward would produce a more stable response than equivalent information alone was found to have certain suggestive support.

References

Bitterman, M. E. Information and effect in incidental learning. *Amer. J. Psychol.*, 1956, **69**, 410–416.

Carr, H. A. The influence of visual guidance in maze learning. *J. exp. Psychol.*, 1921, **4**, 399–417.

Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

Eisenson, J. Confirmation and information in rewards and punishments. *Arch. Psychol., NY*, 1935, Whole No. 181.

Farber, I. E. The spread of effect of reward and punishment in a multiple-choice situation. Unpublished master's thesis, Univer. of Missouri, 1940.

Haslerud, G. M. Anticipative transfer of mechanically guided turns. *J. exp. Psychol.*, 1953, **45**, 431–436.

HILLIX, W. A. Information and reward in human learning. Unpublished master's thesis, Univer. of Missouri, 1954.

HOVLAND, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication*. Princeton: Princeton Univer. Press, 1949.

MARX, M. H. An experimental study of response repetition and success gradients in multiple-choice learning. Unpublished master's thesis, Washington Univer., 1941.

MUENZINGER, K. F. Motivation in learning. II. The function of electric shock for right and wrong responses in human subjects. *J. exp. Psychol.*, 1934, **17**, 439–443.

PORTER, L. W. The effect of "right" in a modified Thorndikian situation. *Amer. J. Psychol.*, 1957, **70**, 219–226.

POSTMAN, L., & ADAMS, P. A. Performance variables in the experimental analysis of the law of effect. *Amer. J. Psychol.*, 1954, **67**, 612–631.

SIMS, V. M. The relative influence of two types of motivation on improvement. *J. educ. Psychol.*, 1928, **19**, 480–484.

STONE, G. R. The effect of negative incentives in serial learning: II. Incentive intensity and response variability. *J. gen. Psychol.*, 1950, **42**, 179–224.

THORNDIKE, E. L. A theory of the action of the after-effects of a connection upon it. *Psychol. Rev.*, 1933, **40**, 434–439.

THORNDIKE, E. L. The direct action of rewards upon mental connections and their indirect action via the stimulation of inner equivalents of the connections. *J. exp. Psychol.*, 1935, **18**, 91–96.

TOLMAN, E. C., HALL, C. S., & BRETNALL, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, **15**, 601–614.

TUCKMAN, J. The influence of varying amounts of punishment on mental connections. *Teach. Coll. Contrib. Educ.*, 1933, No. 590.

WATERS, R. H., & ELLIS, A. J. The relative efficacy of free and guided learning when equated in terms of time. *J. comp. Psychol.*, 1931, **12**, 263–277.

(Received August 10, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 103–110

FREQUENCY AND USEFULNESS OF VERBAL AND NON-VERBAL METHODS IN THE LEARNING AND TRANSFER OF A PAIRED-ASSOCIATE SERIAL MOTOR TASK[1]

EVA NEUMANN

Elmira College

Does the use of verbal cues facilitate the learning, retention, and transfer of a motor task? This problem is both controversial and of recent interest (Battig, 1954; Braun & Bendig, 1957; McAllister, 1953; Rossman & Goss, 1951). Battig (1954, p. 371) has attributed to Brogden the "verbal hypothesis" that "the acquisition of a motor skill will be facilitated to the extent to which verbal descriptions of the stimulus, the response, or both, can be utilized. This hypothesis would predict that if verbal cues are added to the standard test superior acquisition will result." Without adding extrinsic verbal cues to a motor task, Warden (1924) had previously formulated the similar principle that for human mazes the use of a "verbal mode of attack" was superior to "motor" or "visual" methods.

Opposed to these principles are the results obtained by Neumann and Ammons (1958) with the so-called "Schlosberg Board Task," a paired-associate serial motor task devised by Van Dusen and Schlosberg (1948). Neumann and Ammons found no differences in performance associated with different kinds of verbalization used by *Ss*, according to their own choice, either during the acquisition or during the retention periods investigated.

In order to help clarify these apparent contradictions, the present study was designed to investigate the utility of different kinds of verbal cues for learning and transfer of a serial motor task. The strategy of this investigation is to establish four categories or so-called "methods of learning," based on the kind of verbal symbolism employed by *Ss* in the learning of and transfer to different patterns of the Schlosberg Board Task, and to study (*a*) the frequency distribution of these methods, (*b*) their relative learning efficiency, and (*c*) their relative transfer value when different patterns are learned.

METHOD

Subjects

Sixty male undergraduate and graduate students at the University of Kentucky served as *Ss*; they were volunteers and not familiar with the task.

Apparatus

The apparatus used in the present study consisted of three boards constructed to the specifications of the so-called "Schlosberg Board" (Van Dusen & Schlosberg, 1948). Eight pairs of toggle switches were mounted

[1] This paper is taken from a thesis submitted to the Department of Psychology, University of Kentucky, in partial fulfillment of the requirements for the PhD degree. The author wishes to express her appreciation to James S. Calvin and Richard L. Blanton for assistance and guidance. This manuscript was prepared under the sponsorship of the National Science Foundation Research Participation Program held at the University of Michigan in the summer of 1959; the author is indebted to William L. Hays of the University of Michigan for helpful editorial suggestions.

on boards 40 × 40 in. in size and painted dull black. The switches, eight "stimulus switches" and eight "response switches," were arranged on each board in different geometrical patterns, as shown schematically in Fig. 1. The first arrangement of switches was linear and regular (Pattern LIN), the second arrangement was two-dimensional and regular (Pattern CIR), and the third pattern was two-dimensional and irregular (Pattern IRR). Pattern CIR was a replication of the original Schlosberg Board, while Patterns LIN and IRR were variations constructed for the present study. Pattern LIN was expected, on the basis of findings from maze studies (Husband, 1931; Scott, 1930), to give rise to an increase in verbal symbols as compared with Pattern CIR, and Pattern IRR was expected to yield an increase in motor learning. The stimulus switches were connected by a red line, and the starting switch was marked with a red cross. On all boards, each stimulus switch was wired to one and only one response switch in the following randomly determined order: 1–7, 2–4, 3–8, 4–1, 5–6, 7–5, 8–3. The numbers, which were, of course, not visible to *S*, were assigned to the stimulus switches in the order indicated by the red line beginning with the starting switch; the response switches were numbered correspondingly, that is, in Patterns LIN and IRR, a response switch received the same number as the stimulus switch directly above it; in Pattern CIR, it was given the same number as the stimulus switch on the same radius. In addition to the eight pairs of switches, the electrical circuit included a transformer and a buzzer. The buzzer sounded each time the closing of a stimulus switch was followed by the closing of its appropriate response switch. The boards were placed on an upright easel, and *S* was seated in front of the board. A disc recorder was used to record *S*'s verbal responses during the rest periods following each trial.

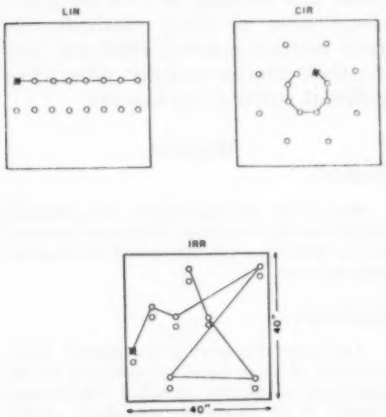

FIG. 1. The three patterns used.

Design

Twenty *Ss* each learned as Task I one of the three patterns (LIN, CIR, or IRR) to a criterion of two successive errorless trials. Immediately after reaching this criterion of learning, they learned one of the other two patterns as Task II. Ten *Ss* were randomly assigned to each of six experimental groups. Pattern LIN was learned as Task I by Groups 1 and 2, and as Task II by Groups 3 and 5; Pattern CIR was learned as Task I by Groups 3 and 4, and as Task II by Groups 2 and 6; Pattern IRR was learned as Task I by Groups 5 and 6, and as Task II by Group 2 and 4. Therefore, each pattern was learned as Task I by 20 *Ss*, and as Task II by 20 different *Ss*; for each pattern of Task II, 10 *Ss* had previously learned one different pattern, and the other 10 *Ss* had previously learned another different pattern.

Procedure

The instructions directed *S* to turn on the starting switch, located on the red line and marked with a cross, and then to try to match this switch by a switch not located on the line. If the pairing had been correct, a buzzer would sound. Whether or not the buzzer sounded, *S* was to proceed to the next switch along the red line, and to attempt to match this switch correctly. He had to continue along the line until he had attempted to match all eight stimulus switches. These eight pairings constituted one trial. After each trial, *S* was told how many seconds he had taken to go through the series. Between trials, *S* turned his back to the board and rested for 30 sec. During the rest periods, he was asked to comment on "the methods he used to learn and remember the task." A relatively free questioning procedure was used to obtain a maximum of classifiable information from *S*, without exerting an undue pressure on him, and without asking leading questions.

Classification of Ss' Verbal Reports

The transcribed records of *Ss*' verbal responses were classified into four categories,

based on the kind of verbal symbolism used to identify the individual switches. The following categories were employed: (*a*) numbering or counting (Method C), (*b*) verbal spatial designation (Method D), (*c*) reference to movement patterns and directions (Method M), and (*d*) reference to visual images (Method V). The classification was based on the predominant method used by *S* during the main period of each task; that is, excluding the first and last trials. During the first trials, *S*s were often not yet able to comment clearly on any definite method; during the last trials, performance tended to have become automatic, and some *S*s made comments resembling references to movements, although they had actually used a different method to learn new combinations. The present classification does not include categories for so-called "mixed methods." The decision about which method is "predominant" admittedly entails a certain degree of arbitrariness on the part of *E*; to minimize this effect, the records were independently rated by a second judge.[2] The original agreement between the judges was 80%; after a common detailed review of all records, the judges agreed on all but one protocol. The following rules for classification were applied:

Counting (*Method C*).—The *S* designates both the stimulus and the response switches by numbers, usually by cardinal numbers. Example: *S* says or reports saying to himself "1 goes with 7," or "1–7." This numbering has to be applied to the entire series of switches on at least one trial.

Verbal spatial designation (*Method D*).—The *S* designates switches by describing their spatial location on the board. Example: "The lower-lower right-hand switch on the inner circle goes with the upper-upper left-hand switch on the outer circle." The stimulus switches may be designated by numbers, usually by ordinal numbers.

Movement patterns (*Method M*).—The *S* designates the response switches not by uniquely determining specific verbal symbols, but refers to them more vaguely in directional terms, frequently supporting his words by pointing at the (imagined) items. Example: "This one goes up here, then back to the center, then down there." The stimulus switches may be designated by ordinal numbers.

Visualization (*Method V*).—The *S* does not designate the response switches at all, but comments on the building up of a so-called "visual image."

Results

Difficulty of the Three Patterns

It was found that Patterns LIN and IRR did not differ from Pattern CIR (the original Schlosberg Board) with respect to learning difficulty as measured in terms of trials to criterion, total time to criterion, or rate of learning expressed as trials to successive criteria (Melton, 1936). All patterns were learned with a grand mean of 23 trials as Task I and a grand mean of 15 trials as Task II. The conditions used in the present experiment (distributed, unpaced learning; 30-sec. rest periods between trials) apparently facilitated learning as compared with those in the Neumann and Ammons study (1958) which required 40 trials with paced learning and 1-min. rest periods, or with the conditions of Van Dusen and Schlosberg (1948) where learning was continuous and accompanied by the enunciation of nonsense syllables, yielding a mean of 31 trials.

Frequency of Methods of Learning

It can be seen from Table 1 that in both Task I and Task II, Methods C and M were more frequently used than Methods V and D. Chi square tests, based on an expected frequency of N/4, showed that the frequencies for the four methods deviated from chance; for Task I, the difference was significant at the 1% level, and for Task II at the 5% level. More *S*s used Method C in Task II than in Task I, and fewer *S*s used Method M in Task II than in Task I. These changes did not, however, make the distribution of methods in Task II significantly different from Task I,

[2] Jacqueline Goodnow of Walter Reed Army Institute of Research, Washington, D. C.

TABLE 1

FREQUENCIES OF METHODS OF LEARNING USED WITH THE THREE PATTERNS IN TASKS I AND II

Pattern	Methods in Task I				Methods in Task II			
	Counting	Motor	Visual	D (Verbal Spatial)	Counting	Motor	Visual	D (Verbal Spatial)
LINear	13	3	1	3	11	3	3	3
CIRcular	3	11	3	3	7	8	3	2
IRRegular	2	15	1	2	6	10	3	1

as was determined by a chi square test.

Within each task, the choice of a method was influenced by the geometrical arrangement of the three patterns. Considering the data for the more important and predictable Methods C and M, it can be seen from Table 1 that Method C was used most frequently in Pattern LIN and less frequently in Patterns CIR and IRR. The frequency at which Method M was used, however, increased systematically from Pattern LIN to Pattern CIR, and from Pattern CIR to Pattern IRR. In other words, Pattern LIN was found to be conducive to Method C, and Pattern IRR was found to be conducive to Method M. This result is in agreement with expectation based on the findings from mazes (Husband, 1931; Scott, 1930).

Pattern CIR elicited in Task I a greater frequency in the use of Method M than in the use of Method C, but in Task II the Methods C and M were used with about equal frequency. It is interesting to note that in the previous experiment with the Schlosberg Board (Neumann & Ammons, 1958) the distribution of methods was more similar to the distribution of methods in Task II of the present study than to the distribution obtained in Task I. Apparently, the more stressful experimental conditions in the previous study, such as pacing by a metronome, and the longer rest periods between trials, prompted more *Ss* to adopt the more precise numerical method early in the learning process, whereas the less stressful conditions in the present study and the shorter rest periods were more conducive to the use of the more direct and "primitive" Method M during Task I; only the greater familiarity with the experiment in Task II enabled *Ss* to use Method C more often.

The difference in frequency of methods was found by means of chi square tests to be significant at the 1% level for the three patterns in Task I, and at the 5% level for the three patterns in Task II. In Task I, Pattern LIN differed significantly at the 1% level from Patterns CIR and IRR with respect to the frequency of methods, but these differences were not significant in Task II.

Consistency in Method

Analysis of the methods chosen by each *S* in Task I and Task II indicated the existence of two further tendencies beyond the general shift toward Method C in Task II: (*a*) a tendency for consistency of method, regardless of the fact that a different method might be more adequate for the different pattern in Task II, and (*b*) a tendency to employ a method

more appropriate to the geometrical nature of the pattern in Task II. Inspection of Table 2 shows that the tendency for consistency is the stronger one; 35 out of 60 *Ss* used the same method in both tasks. This concordance was found to be significant at the 1% level by means of a z score based on the comparison of obtained frequencies in the diagonal cells with frequencies expected to be in the diagonal cells of the table; the expectation of what method would be used in Task II was based on the information about what method each *S* had used in Task I.

Evidence for the weaker tendency, namely, to adopt a more adequate method in Task II, can be seen in the fact that all of the 11 changes in method from Task I to Task II which involved Methods C and M, took place in the predicted direction. When the *Ss* were shifted from Patterns CIR or IRR to Pattern LIN, 8 *Ss* changed their method from M to C; when shifted from Pattern LIN to Pattern IRR, 3 *Ss* changed from Method C to Method M. No changes involving Methods C and M occurred between Patterns CIR and IRR, which were both relatively conducive to motor learning, or from Pattern LIN to Pattern CIR. Apparently, having used the counting method appropriate to Pattern LIN first, made its use easier for Task II. But when two patterns were more drastically different, such as Patterns LIN and IRR, changes were more likely to occur.

TABLE 2

FREQUENCIES OF METHODS USED BY THE SAME *Ss* IN TASK I AND TASK II FOR ALL CONDITIONS COMBINED

Task I	Task II				
	C	M	V	D	Total
C	*14*	3	1	0	18
M	8	*15*	4	2	29
V	0	1	*3*	1	5
D	2	2	1	*3*	8
	24	21	9	6	60

Relative Learning Efficiency of the Methods Employed

The use of any particular method of learning employed in Task I of the present study was not associated with greater learning efficiency than the use of the other methods.

The mean number of trials to criterion differed very little as a function of the method employed by *S*. The means were 22.1 trials for Method C, 23.3 trials for Method M, 24.2 trials for Method V, and 25.8 trials for Method D. A Kruskal-Wallis test for mean trials to criterion for the four methods, regardless of pattern, failed to indicate a significant difference between methods. No significant difference was found, by means of the same test, when the means for trials to criterion were compared for the methods within each pattern.

The mean total time required to learn Task I to criterion was 761 sec.; no significant difference was shown by Kruskal-Wallis tests to exist between methods regardless of pattern or between methods within patterns.

Curves representing the rate of learning plotted in terms of trials or time to criterion (Melton, 1936) showed a very similar time course for the four methods of learning; particularly, the curves for Methods C and M run very close to one another. The learning curve for Pattern LIN is shown as a sample in Fig. 2.

The results justify the conclusion that, for the first learning of this serial motor task, learning was not

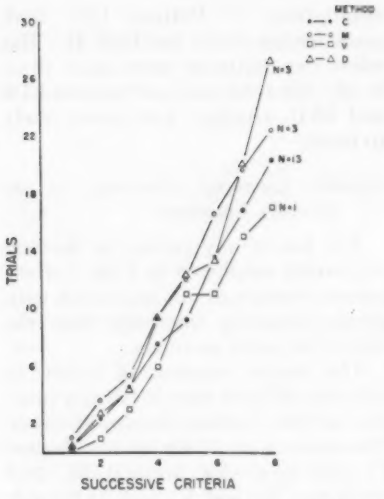

FIG. 2. Trials to successive criteria on Task I, Pattern LIN.

facilitated by the use of verbal or numerical cues. This finding apparently contradicts the generality of the "verbal hypothesis" (Battig, 1954) which states the expected superiority of verbal cues for the learning of motor tasks.

Transfer Effects

The *S*s who had used Method C in Task I performed better on the first trials of Task II than *S*s who had used the other three methods in Task I. The *S*s who had used Method C in Task I achieved means of 4.0 correct responses on the first transfer trial and 14.1 correct responses on the first three transfer trials combined, as compared with the respective means of 1.5 and 6.2 correct responses in Task II resulting from the use of the other three methods in Task I combined. These differences were found, by means of an analysis of variance, to be significant at the 1% level.

This result is in agreement with the theoretical consideration that the use of Method C would lead to the recognition of the same numerical relationship between the stimulus and response switches in Task I and Task II, and would therefore be associated with the best performance on the early transfer trials.

In terms of trials required to learn Task II to criterion as well as in terms of difference scores derived by subtracting the number of trials to criterion in Task II from the trials in Task I, learning of Task II was considerably facilitated by the preceding use of verbal as well as of numerical cues in Task I. The *S*s who had used Method C required a mean of 10.6 trials to learn Task II to criterion, and *S*s who had used Method D in Task I took 12.8 trials to learn Task II. After using Methods M and V in Task I, 16.2 and 22.3 trials, respectively, were necessary to learn Task II. The differences among the means were found by means of the Kruskal-Wallis test to be significant at the 5% level; when the combined effects of Methods C and D were contrasted by the same test with the combined effects of Methods M and V, the difference was significant at the 1% level.

The difference scores as a function of the method used in Task I were as follows: 14.0 trials were saved in Task II after the use of Method C and 11.8 trials after the use of Method D, whereas Method M yielded only a saving of 5.7 trials and Method V a saving of 2.8 trials. The differences among the mean difference scores were found, by means of the Kruskal-Wallis test, to be significant at the 1% level.

While it was expected that the previous use of Method C would facilitate the learning of Task II, the

facilitating effect of Method D was quite surprising. Method D appeared to be rather circumstantial and awkward, and seemed rather uneconomical for the learning of Task II. Apparently the great precision of this method became more useful with greater familiarity with the experimental situation. It can be concluded from the transfer data in the present study that the learning of the second pattern (Task II) is greatly facilitated by the preceding use of the numerical or verbal method in Task I.

Discussion

The results of the present study obviously represent some contradiction to "verbal hypothesis" (Battig, 1954), as well as to Warden's (1924) principle of the superiority of verbalization for maze learning. The present task was learned equally well by numbering, verbal spatial designation of items, references to movements, and visualization. It should be noted that previous *Es* such as Warden (1924) and Husband (1931) failed to differentiate between numerical and purely verbal cues, whereas important differences can be found if the effects of these methods on particular phases of the learning process are examined.

Brogden may have considered mainly single-response or continuous motor tasks, and these may very well fall in the domain of the principle stated. However, for serial or multiple-response motor tasks, it seems very doubtful that this principle holds without qualification. The results of the present study should caution against too early theoretical generalizations for all motor tasks; it seems that more research is needed to determine the generalizability of laws from particular kinds of motor tasks to all motor tasks.

The generality of the principle seems to be also questionable in the light of the present finding that even in those cases in which *Ss* choose a method presumably most adequate for a given geometrical pattern, the use of that method was not associated with greater ease of learning the pattern.

The superiority of verbalization for transfer in the present study should also not be taken as evidence for the principle; it might very well be a consequence of the particular conditions in this experiment. Verbalization probably was more helpful in detecting the "identical element," namely, the same numerical connection between stimulus and response switches in both tasks, whereas the movements leading to correct responses or the visual image of the correct solution had to be learned "from scratch" in Task II when methods other than complete verbalization were used.

Summary

This experiment was designed to test the hypothesis that the use of verbal cues facilitates the learning of motor tasks. It investigated the role of different kinds of verbalization in the learning and transfer of a paired-associate serial motor task, the so-called Schlosberg Board Task. The 60 *Ss* learned two out of three different spatial arrangements of the task and reported on their method of learning. Four methods of learning were distinguished: Method C (counting), Method D (verbal spatial designation), Method M (motor references) and Method V (visual image). The geometrical nature of a pattern influenced the choice of a method of learning. The use of verbalization did not facilitate the original learning of the task, but transfer was considerably facilitated by the use of verbal cues.

References

Battig, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a lever-positioning skill. *J. exp. Psychol.*, 1954, **47**, 371–380.

Braun, H. W., & Bendig, A. W. Effect of addition of irrelevant verbal cues on perceptual-motor learning. *J. exp. Psychol.*, 1957, **54**, 105–108.

Husband, R. W. Comparative behavior on different types of mazes. *J. gen. Psychol.*, 1931, **5**, 234–243.

McAllister, D. E. The effects of various kinds of relevant verbal pretraining on subsequent motor performance. *J. exp. Psychol.*, 1953, **46**, 329–336.

MELTON, A. W. The end spurt in memorization curves as an artifact of the averaging of individual curves. *Psychol. Monogr.*, 1936, **47,** (2, Whole No. 212).

NEUMANN, E., & AMMONS, R. B. Acquisition and long-term retention of a simple serial perceptual-motor skill. *J. exp. Psychol.*, 1958, **53,** 159–161.

ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: The role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42,** 173–182.

SCOTT, T. C. The retention and recognition of patterns in maze learning. *J. exp. Psychol.*, 1930, **13,** 164–207.

VAN DUSEN, F., & SCHLOSBERG, H. Further study of the retention of verbal and motor skills. *J. exp. Psychol.*, 1948, **38,** 526–543.

WARDEN, C. J. The relative economy of various modes of attack in the mastery of the stylus maze. *J. exp. Psychol.*, 1924, **7,** 243–275.

(Received August 12, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 111-120

LIKELIHOOD JUDGMENTS AND SEQUENTIAL EFFECTS IN A TWO-CHOICE PROBABILITY LEARNING SITUATION [1]

NORMAN H. ANDERSON [2] AND RICHARD E. WHALEN [3]

Yale University

Despite some rather remarkable successes of statistical learning theory models for two-choice probability learning situations (Burke, Estes, & Hellyer, 1954; Estes & Straughan, 1954; Schoeffler, 1954), it has become apparent that there is much in *Ss*' behavior in this situation which is not accounted for by the models. The best known discrepancy is the "gambler's fallacy," or negative recency effect, the term applied to the decrease in predictions of a given event over the later part of a run of consecutive occurrences of that event. This phenomenon, first reported formally by Jarvik (1951), has been studied more extensively by Nicks (1959) and by Anderson (1960). The latter report made a detailed investigation of sequential dependencies and found other discrepancies between the models and the behavior. It was therefore suggested that the observed agreement with the event "matching" behavior predicted by the models does not confirm the theoretical assumptions made in the models since this agreement apparently results from an averaging out of sequential increments and decrements in response probability which are not accounted for by the models.

The present experiment had two purposes: first, to continue the parametric study of sequential dependencies, and second, to explore the use of another response measure which might be more revealing than frequency of predictions. To accomplish these ends, *Ss* were asked not only to predict one of two events, but at the same time to indicate the likelihood of the event predicted.

METHOD

Subjects.—The *Ss* were 54 student nurses from the Grace-New Haven hospital and the Yale School of Nursing. Each received $1.50 for her services.

Apparatus and procedure.—The *Ss*, who were tested individually, were seated in front of a 4 × 3-ft. Masonite panel mounted on a desk. The panel contained a 16½-in. row of 11 equally spaced ¼-in. holes approximately 10 in. above the base. A shielded signal light was located 6 in. above the row of holes, and a 6-w. opal lamp, the reinforcing stimulus, was placed 2 in. from either end of the row of holes. The *S* indicated her response by placing a pointed plastic stylus in one of the 11 holes according to the instructions described below. The *E* sat behind the panel, recorded *S*'s responses, and flashed the appropriate reinforcing light on each trial.

The trial duration, 10 sec., was controlled by a synchronous motor and cam which lit the signal light for 2 sec. during which time *S* was to make her response. Approximately 1 sec. after the signal light went off, *E* flashed the scheduled reinforcing light for about 1 sec. The *S* was instructed to leave the stylus in the hole until the reinforcing light had flashed.

Each of the 11 response holes in the board was labeled on *S*'s side. The center hole was labeled Equally Likely. Extending outward in both directions from the center hole the labels read Slightly More Likely, Somewhat More Likely, Quite a Bit More Likely,

[1] This research was supported by a grant from the George Sheffield Scientific Fund. The authors wish to thank A. Parducci for a critical reading of the manuscript.

[2] Now at the University of California, Los Angeles.

[3] Now at St. Elizabeths Hospital, Washington, D. C.

Very Much More Likely, and Almost Certain, the last being the label over the two end holes.

Each *S* was read instructions which stated that the task had two parts. The first part was to predict which light would come on next, and it was emphasized that all of the five holes on the side of the predicted light were equivalent in this respect. The second and more important part of the task was to indicate *at the same time* how likely it was that the predicted light would come on by choosing the hole with the appropriate label. The center hole, of course, was to be used when it was thought that neither light was more likely to flash than the other. In addition, *S* was told that the way the lights came on was in part a matter of chance so that it was not possible to be correct on every trial, but that it was possible to improve with practice.

Eight practice trials, on which both reinforcing lights were flashed, were then given. The *S* was then asked to repeat the directions and give the meaning of the possible responses. Following this there were 350 trials without conversation or interruption.

Design.—In the three main reinforcement conditions the frequencies of the two reinforcing lights were 50:50, 65:35, and 80:20. Within each of these conditions, an orthogonal $2 \times 2 \times 3$ design was formed with two *Es*, two positions (right and left) of the more frequent light, and three random sequences of reinforcing lights as factors. The first 50 trials of each sequence were random, subject to balancing according to the specified frequencies. Three 100-trial blocks were also constructed for each of the three frequency conditions using a similar balancing restriction for each 100-trial block. Within *each* condition of the three-way design, these three blocks were used in a 3×3 latin square arrangement such that each row of the square formed the last 300 trials of one of the sequences in that condition.

It will be noted that with such extensive counterbalancing, there would be very few error degrees of freedom within a given condition. However, on the basis of previous experience (Anderson, 1960; Anderson & Grant, 1957), it was decided that certain interactions, mostly those involving position, would be negligible. These interactions were not evaluated but instead were automatically pooled with the error term. This a priori pooling procedure has certain dangers but, when it is based on reasonable evidence, it is probably preferable to a posteriori pooling on the basis of nonsignificant *F*s. In any case, this technique of pooling is superior to randomizing the effects of the counterbalanced variables as well as to counterbalancing without evaluation of even the main effects.

Results

For purposes of analysis, the response holes were numbered from +1 to +5 on the preferred (E_1) side of the response panel, and from −1 to −5 on the nonpreferred (E_2) side. The center hole was numbered 0. The right-hand side was arbitrarily taken as the preferred side for Group 50. For the other two groups the preferred side was the side of the more frequently occurring reinforcing light, E_1. Two main dependent variables were considered. A frequency score was obtained by adding to the number of E_1 responses one-half the number of center hole responses. A magnitude score was obtained by taking the algebraic mean of the likelihood ratings, using the above numbering system, with due regard to sign.

Frequency scores.—Mean frequencies of the 11 responses over trials 251–350 are shown in Fig. 1. The regular ordering of the three groups on the nonpreferred side is as would be expected; Group 80 makes fewer responses on this side than Group 65, etc. However, on the preferred side the only marked difference between groups is in the +5 response frequency. In the middle, the extreme dip indicates that the center hole was chosen quite infrequently. Surprisingly, there is little difference between groups in this respect. Finally, it should be noted that Groups 65 and 80 show a U-shaped distribution of preferred side responses, the +2, +3, and +4 responses all being less frequent than the +1 or +5 responses. This result was unexpected as it had been thought that the curves would either be monotone or resemble

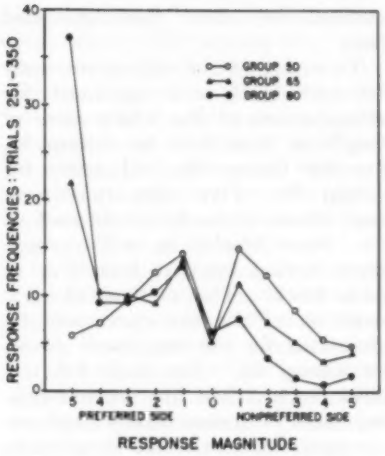

FIG. 1. Mean frequencies of the 11 response categories for the three reinforcement schedules over Trials 251–350. Preferred side responses are predictions of the more frequent light, E_1.

an inverted U. A *post hoc* analysis of the +1 and +2 response data for these two groups, ignoring the other factors in the design, indicated that +2 responses were significantly less frequent than +1 responses ($F = 5.06$, $df = 1/34$).

In order to compare the present results with those obtained in the standard two-choice probability learning situation, the frequency scores are given in Table 1a as a function of groups and trials. The analysis in Table 2 shows that both of these variables had a significant effect. Since Group 50, as would be expected, changed little over trials, there should be a differential trend over trials for the three groups. The most sensitive test for this differential trend is the Linear × Linear component of the T × G interaction, and this is seen to be significant, although the over-all interaction is not. The mean frequency scores over trials 251–350 were not significantly different from the "matching" behavior predictions of 50, 65, and 80% for the three respective Groups. To this extent, at least, the present technique yields data comparable to those obtained when only two choices are permitted.

Magnitude scores.—There are several ways to treat the likelihood judgment data. Table 1b shows the response magnitudes averaged over just those trials on which a preferred side response was made. These data are thus "corrected" for response frequency. The analysis of these data, given in Table 2, parallels the analysis of the frequency scores. Similar trends

TABLE 1

FREQUENCY SCORES AND LIKELIHOOD JUDGMENTS AS A FUNCTION OF TRIALS AND GROUPS

Group	Trials		
	51–150	151–250	251–350
a. Frequency Scores			
50	50.1	51.3	51.8
65	58.7	62.4	67.1
80	73.7	79.2	82.1
b. Likelihood Judgments: Preferred Side Responses			
50	2.36	2.41	2.42
65	2.86	3.07	3.14
80	3.01	3.18	3.45
c. Likelihood Judgments Disregarding Sign: All Trials			
50	2.35	2.41	2.36
65	2.65	2.76	2.83
80	2.67	2.86	3.14
d. Magnitude Scores			
50	0.01	0.10	0.15
65	0.70	1.07	1.39
80	1.79	2.18	2.54

Note.—See text for definition of scores.

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL FREQUENCY SCORES (TABLE 1a) AND OF MEAN "CORRECTED" LIKELIHOOD SCORES (TABLE 1b): TRIALS 251–350

Source	*df*	*F* Ratios	
		Frequency	Magnitude
Between			
Groups	2	100.36*	6.87*
Es	1	.21	.05
Position	1	.86	1.75
G × *E*	2	.15	1.31
G × P	2	.96	3.57
Sequences (pooled)	6	.27	.77
Error (b)	39	(100.51)	(1.42)
Within			
Trials	2	22.81*	8.82*
T × G	4	3.13	1.74
T × G: linear × linear	1	8.81*	6.10*
T × P	2	.11	1.25
T × *E*	2	.23	.04
Blocks (pooled)	6	1.40	3.67*
B × P (pooled)	6	2.67*	1.53
Square residual (pooled)	6	.86	1.69
Error (w)	78	(22.89)	(0.103)

Note.—Error mean squares in parentheses.
* $P < .05$.

are found when response magnitudes, disregarding sign, are averaged over each set of 100 trials as shown in Table 1c. Finally, the entries in Table 1d are the mean magnitude scores computed with due regard to sign as defined above. The trends over trials and groups reflect changes in response frequency as well as in the size of the likelihood judgments so that the differences are more marked than in Tables 1b and 1c. These magnitude scores should thus give maximal sensitivity to the effects of the experimental variables, although they were not analyzed here.

Sequential dependencies.—The main results of the experiment are the subsequence curves, i.e., the successive responses given during the course of some designated type of subsequence of reinforcing lights. Previous results suggested that the most informative of the subsequence curves would be the run and alternation curves, and these were examined here.

To construct the run curves, each 100-trial block was examined for subsequences of E_1s which were of length at least four for Group 50, five for Group 65, and seven for Group 80. Five such runs were were chosen at random from each of the three blocks in each group, there having been at least five in each block. The method of construction of the run curve will be illustrated for the magnitude scores for Group 65. For each 100-trial block, each *S* had five ordered subsequences of (consecutive) responses corresponding to the five occurrences of the subsequence of reinforcing lights, $E_2E_1E_1E_1E_1E_1$. The first response used followed the E_2 occurrence and, hence, immediately preceded the first E_1 occurrence. The second, third, etc., responses in the subsequence followed the first, second, etc., occurrences of E_1. The five ordered subsequences of responses were next summed, separately for each ordinal position in the subsequence, to yield a single ordered subsequence for each *S* for each block. These summed subsequences were the unit data in the statistical analysis. The procedure for the magnitude scores for the other groups, and for the frequency scores was identical.

The resulting run curves, divided by five to give trial means, are shown in Fig. 2. The curves in the three successive panels for each group are based on comparable stimuli because of the latin square arrangement of the 100-trial blocks used in the sequences. The frequency and magnitude curves are quite similar in shape. In every case the curve rises over the early trials of the run, showing a positive effect of the reinforcing stimulus. In every case there

is a downward trend over the later trials which demonstrates the "gambler's fallacy," or negative recency effect. The curves are flatter and peak later in the subsequence for the more extreme reinforcing schedules. In addition, there is some tendency for the peaks of the curves to appear later in the subsequence in the later trials for Groups 65 and 80.

On the basis of previous data, the a priori decision was made to test for the negative recency effect by comparing the responses at abscissa values 2 and 4 for Group 50 and at abscissa values 3 and 5 for Group 65. The resulting analyses are thus split-plot, or mixed, designs (see Lindquist, 1953; Type IV and VI designs), and are given in Table 3. For the most part, the analyses of frequency and magnitude yield comparable results and no distinction between them will be made in what follows. However, it should be noted that the significant F ratios tend to be larger for the magnitude scores. This suggests that the power of tests based on magnitude scores will be greater than for tests using frequency scores or, in other words, that the magnitude score is the more sensitive to the effects of the experimental variables.

The F ratios for Ordinal Position in Run show that the responses being compared are in each case significantly different, and hence that there is a real negative recency effect. The significant Trials effect for Group 65 merely indicates that the latter half of the run curve is higher in the later trials for this condition.

It had been planned to treat the Group 80 run curves in the same

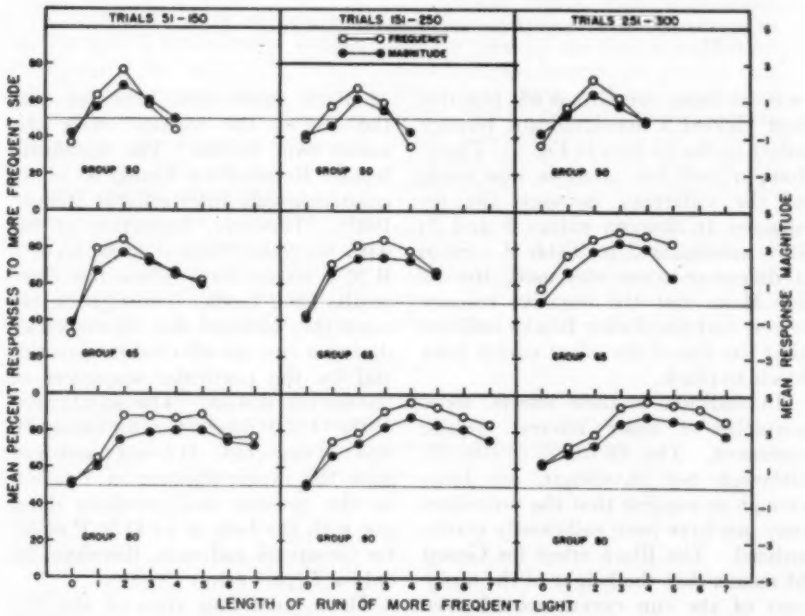

FIG. 2. Run curves for frequency and magnitude scores as a function of reinforcement schedule and trials.

TABLE 3

ANALYSIS OF VARIANCE OF SELECTED POINTS ON RUN CURVES

Source	*df*	Group 50		Group 65		Group 80	
		Freq.	Magn.	Freq.	Magn.	Freq.	Magn.
Between *Ss*							
Mean	1					11.84*	13.38*
Es	1	.75	1.82	2.31	3.90	1.31	3.28
Position	1	.68	.11	.68	.02	.02	.22
Sequences	2	.60	.13	1.78	.57	1.98	2.56
Error (b)	13	(4.92)	(180.25)	(97.66)	(98.08)	(1.27)	(58.74)
Within *Ss*							
Trials	2	1.96	2.46	9.16*	8.75*	.07	.08
T × *E*	2	.79	.80	1.24	.92	.72	1.75
Blocks	2	.26	1.38	22.26*	24.64*	2.37	5.02*
B × P	2	.57	1.92	.61	.48	2.71	1.53
Square residual	2	.18	.35	2.43	6.85*	.73	1.33
$Error_1$ (w)	26	(1.41)	(48.41)	(75.28)	(36.81)	(1.39)	(46.27)
Ordinal position in run	1	35.70*	75.26*	17.01*	22.15*		
O × P	1	3.42	22.38*	1.68	1.53		
O × *E*	1	.05	.00	.51	.11		
$Error_2$ (w)	15	(1.66)	(25.36)	(57.15)	(37.87)		
O × B	2	.60	.80	2.42	3.56*		
O × T	2	.94	.20	.75	.36		
$Error_3$ (w)	32	(0.50)	(27.94)	(86.23)	(28.12)		

Note.—Cell entries are *F*s except for error mean squares in parentheses.
* $P < .05$.

way as those for Group 65, but this test yielded a nonsignificant recency effect as can be seen in Fig. 2. Therefore, a *post hoc* analysis was made of the difference between the responses at abscissa values 5 and 7; it is summarized in Table 3. Since a difference score was used, the *F*s for Mean test the negative recency effect, and the *F*s for Blocks indicate that the size of the effect varied from block to block.

In addition to these effects, some variables of lesser interest require comment. The *F*s for *E* (Table 3), although not significant, are large enough to suggest that the procedure may not have been sufficiently standardized. The Block effect for Group 65 means that the height of the latter part of the run curves varied over the three 100-trial blocks. For this same group, the Ordinal Position × Block interaction indicates that the size of the recency effect also varies over blocks. The significant Square Residual for Group 65 is not unambiguously interpretable (Grant, 1948). However, inspection of the data suggested that it arose from a B × T interaction. These last three results need further investigation because they indicate that the responses during a run are affected by practice and by the particular sequences of reinforcing stimuli. The significance of the O × P interaction for Group 50 was not expected. It is not consistent with the nonsignificance of Position in the present and previous data, nor with the lack of an O × P effect for Group 65 and may, therefore, be only a Type I error.

There were four runs of the *less* frequent light for Group 65 over Trials 51–350 which were of length

at least four. The data from these runs yielded means, comparable to the first five points of the Group 65 curves of Fig. 2, of 2.1, −0.1, −1.1, −1.0, −0.9 for magnitude, and 68, 44, 37, 37, 37 for frequency. The shapes of these curves, when reflected about the chance level, are not dissimilar to the Fig. 2 curves for the *more* frequent light although the over-all elevation is less. However, it should be noted that they show little or no negative recency effect.

Data analogous to the run curves for subsequences which embodied alternations are given in Table 4. The entries are trial means based on all instances over Trials 51–350 of subsequences of the alternation type as specified in the table. There were 18 such instances for Group 50, and 8 for Group 65. The data show an extreme shift in response following the first occurrence of the less frequent light, E_2, and a consistently smaller shift following the second E_2 occurrence. The frequency scores following the first and second occurrences of E_2 were significantly different ($F = 7.72$ and 8.95, $df = 1/17$, for Groups 50 and 65, respectively).

Since the response following the second E_2 is not greater than 50%, the data of Table 4 do not unequivocally demonstrate the pattern of alternation behavior found by Anderson (1960), and so are perhaps more in accord with the results of Nicks (1959). Although there are a number of procedural differences in the three experiments, it may be suggested that alternation responding is controlled in part by the degree of perceptual "isolation" of the reinforcing stimuli. The stimulus display was several feet in front of *S* in the one experiment which found alternation behavior whereas it was quite close to *S* in the other two experiments. This conjecture is consistent with the contrasting results regarding relative difficulty of learning pure alternation and pure repetition sequences as found by Goodnow and Pettigrew (1956) and by Anderson (1960).

Discussion

Empirical results.—The most easily described characteristic of the shape of the run curves is the trial at which they peak. In this respect, the present data differ from those of Estes (1957) and Nicks (1959). In Nicks' data, the peaks occur after the first E_1 for a 50:50 condition, one trial earlier than for the present data. For Nicks' 67:33 and 75:25 conditions, the peaks occur after the second E_1 which is again an earlier peak than is seen in the roughly comparable 65:35 and 80:20 conditions used here. The most likely source of this discrepancy appeared to be the greater

TABLE 4

Subsequence Curves for Alternation Subsequences

Group and Measure	Stimulus Preceding Tabulated Response						
	X	E_1	E_1	E_2	E_1	E_2	E_1
Group 50							
Frequency	58	64	59	31	56	45	
Magnitude	0.48	0.89	0.74	−1.07	0.40	−0.43	
Group 65							
Frequency	61	68	73	28	73	52	59
Magnitude	0.79	1.57	1.76	−1.07	1.51	0.01	0.83

number of overt responses allowed in this experiment. To test this, the present run curve procedure was used on the Group 50 data of Anderson (1960) in which the standard two-response technique was used. The resulting run curve, based on data of 40 *Ss* over Trials 51–350, was 38, 55, 68, 67, 58, where successive entries are to be compared to successive points of the Group 50 frequency curves of Fig. 2. This curve peaks at the same place in the subsequence but does not show as great a decline on the two following trials as the Fig. 2 curves. The source of the discrepancy between the present data and those of Nicks thus remains unclear. No peak was found in Estes' data, a result which may be due to the small *N*, or to the fact that only runs of length three were considered and these were averaged over a number of different reinforcement schedules. However, it should also be realized that there may be no negative recency effect when, as was the case in Estes' experiment, consecutive short blocks of trials embody different reinforcing schedules and when the sequences are completely random rather than balanced over trial blocks.

Three different procedures have been used in constructing subsequence curves and a brief comment on them may help to avoid confusion. The procedure used by Anderson (1959, 1960) has been discussed in the latter report. It need only be mentioned here that it will be most useful for estimating parameters or testing goodness of fit, but that it may give run and alternation curves of different shapes than the other methods. In Nicks' (1959) procedure, the subsequence $E_2E_1E_1E_1E_1E_1$ contributes to each of the first six points of the curve (Nicks does not actually plot the response following the E_2). However, the subsequence $E_2E_1E_1E_1E_2$ contributes to just the first four points of the same curve. Consequently, successive points on the curves are based on progressively less data, and so are progressively less reliable. This procedure may be useful when data are scarce, but the results are difficult to handle statistically. The procedure used here is apparently the same as that of Estes (1957) and would seem to be optimal for making statistical tests of the raw data, whether formally or by eye.

As far as can be seen at present, obtaining likelihood judgments does not affect the two-choice properties of the task. The "matching" behavior found here is in agreement with classical results (Grant, Hake, & Hornseth, 1951), and the similarity of the frequency and magnitude subsequence curves gives further support to this idea. It would thus appear that the use of likelihood judgments will be valuable in future work since a quantitative response measure should be more revealing of *S*'s response tendencies than a Yes-No score. Nevertheless, further work on the method for obtaining likelihood judgments is needed. It would appear from Fig. 1 that the Equally Likely response can be omitted without injustice to the behavior. It is not clear, however, whether the modes at $+1$ and $+5$ represent an inadequacy of the response scale or whether they reflect two real modal response tendencies, either within or between *Ss*.

Implications.—A simple reinforcement theory would hold that each occurrence of a reinforcing stimulus results in an increase in response probability. As has been noted by previous workers, the negative recency effect seen in the run curves is inconsistent with such a formulation. For the same reason, the negative recency effect is inconsistent with current mathematical models for the two-choice probability learning situation. Most models, excepting perhaps those of Burke and Estes (1957) and Anderson (1960), would also fail to predict the alternation curves in which the response level following the first E_2 is less than that following the second E_2.

Cognitive theories have been popular for the two-choice probability learning situation, but they appear inconsistent with the lack of a sharp negative recency effect in the run curves for the less frequent light for Group 65. A simple

expectancy point of view would presumably explain the fact that the peaks occur later in the run curves of Fig. 2 for the more extreme reinforcing schedules as resulting from expectancies for runs of different lengths which are developed by the reinforcing schedules. From this reasoning, it would follow that *Ss* ought to develop different expectancies for run lengths for the two lights of Group 65. Indeed, since the average run of the less frequent light in Group 65 is shorter even than the average run in Group 50, the run curve for the less frequent light should exhibit a marked negative recency effect, contrary to the observed result. Further investigation is desirable, however, since the amount of data involved in this result is relatively small. Interspersing test runs would yield more data, but a careful design would be needed to check whether the test runs had modified the behavior.

Hake and Hyman (1953) have put forth the view that *Ss* respond to short stretches of the sequence of reinforcing lights in making their responses, and the results of Nicks (1959) and Anderson (1960) support this idea. In the latter report, it was suggested that the behavior might be handled by mathematical models which took immediately preceding subsequences as the discriminative stimuli of a system of S-R relations within the Burke-Estes (1957) formulation of probabilistic discrimination learning. From a cognitive point of view, these S-R relations could be called hypotheses, strategies, expectancies, etc. However, when the overt response is the resultant of several intervening variables or underlying response tendencies, a quantitative treatment is essential to an adequate theory. To date, cognitive approaches have not been provided with the required mathematical apparatus.

As has been noted in earlier reports (Anderson, 1959, 1960), sequential dependencies in the sequence of stimuli and responses give the most direct indication of the action of the various reinforcing events. They will, therefore, be generally more efficient in the description and analysis of behavior than the learning curve. It would seem that, in the initial stages of investigation of complex S-R systems, the subsequence curves will be the most useful of the various sequential dependencies that can be computed.

Summary

In order to extend previous findings concerning sequential dependencies in probability learning tasks, and to test the usefulness of another response measure, 54 student nurses were tested on a multiple-choice version of the standard two-event probability learning problem. The *Ss* not only predicted which one of the two reinforcing events would occur next, but also indicated the likelihood of the event predicted.

Three main groups, which were tested on 50:50, 65:35, and 80:20 reinforcement schedules, showed typical probability "matching" behavior. In all further analyses, the predictions and the likelihood judgments yielded parallel results. It appeared that the likelihood judgments gave a more sensitive response measure but did not change the two-choice character of the task.

The distribution of responses over the likelihood scale was quite regular but showed two peculiarities. The Equally Likely response was used quite infrequently but about equally often in all reinforcement conditions. In addition, the distribution of responses on the half of the scale corresponding to the more frequent event was roughly U-shaped for the 65:35 and 80:20 conditions.

Subsequence curves for runs of a given reinforcing event showed the negative recency effect except for runs of the less frequent event in the 65:35 condition. The peak of the run curves came later in the run for the more extreme schedules. Subsequence curves for alternations in the stimulus event sequences were also studied. The results of the analyses of the subsequence curves were inconsistent with simple S-R and with simple expectancy theories.

References

Anderson, N. H. An analysis of sequential dependencies. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

Anderson, N. H. Effect of first-order conditional probability in a two-choice learn-

ing situation. *J. exp. Psychol.*, 1960, **59**, 73–93.

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

BURKE, C. J., ESTES, W. K., & HELLYER, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.

ESTES, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GOODNOW, J. J., & PETTIGREW, T. F. Some sources of difficulty in solving simple problems. *J. exp. Psychol.*, 1956, **51**, 385–392.

GRANT, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **54**, 64–74.

JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

NICKS, D. C. Prediction of sequential two-choice decision from event runs. *J. exp. Psychol.*, 1959, **57**, 105–114.

SCHOEFFLER, M. S. Probability of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1954, **48**, 323–329.

(Received August 15, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 121–125

WOODWORTH SCALE VALUES OF THE LIGHTFOOT PICTURES OF FACIAL EXPRESSION [1]

NISSIM LEVY

University of Portland

AND HAROLD SCHLOSBERG

Brown University

Engen, Levy, and Schlosberg (1957) described a new series of 48 pictures of facial expressions, posed by a talented actress, Marjorie Lightfoot. There has been considerable demand for this series for use in cross-cultural studies, psychiatric research, and laboratory classes.[2] In 1958 the same authors published norms for the three basic dimensions: Pleasantness- Unpleasantness, Attention-Rejection, and Sleep-Tension. The usefulness of these dimensions in the analysis of facial expressions has been demonstrated (Schlosberg, 1952, 1954; Triandis & Lambert, 1958; Woodworth & Schlosberg, 1954) but conventional category names, like Mirth and Anger, are more appropriate for some research studies. The present paper presents norms in terms of the six categories of the Woodworth Scale comparable to those published by Schlosberg in 1941 for the Frois-Wittmann series. It also considers the effects of differences in exposure time, sex of judges, and methods of calculation on these norms.

[1] Analysis of the data reported in this study was supported in part by USPHS Grant M-2779(A).

[2] The Lightfoot series has been published, in reduced size, in Morgan's manual (1956). The Brown photo lab will supply a set of "jumbo" prints for approximately $5.00, or of a direct print of the 35-mm. negative strip, suitable for mounting in 2 × 2-in. slides, for $2.25. They also have a negative of the original Frois-Wittmann series, consisting of 72 pictures, and will fill similar orders at about $7.50 and $4.00, respectively. Schlosberg has some sets of smaller reproductions of the Frois-Wittmann series, which he will be glad to furnish for research purposes, as long as the supply lasts.

METHOD

The pictures were always presented by projecting 2 × 2-in. slides to groups of 20 to 40 students in a laboratory section of the elementary course at Brown University. Each *S* had a mimeographed blank, with the lines numbered to correspond to the 48 pictures, and the columns labeled in terms of the six Woodworth categories (1. Love, Mirth, Happiness; 2. Surprise; 3. Fear, Suffering; 4. Anger, Determination; 5. Disgust; 6. Contempt). The numbers used to identify the pictures were those that were originally assigned at random, and appeared on the projected pictures. There was enough light in the room to permit each student to record his judgment by a check in the appropriate column, but the dim light did not obscure the brilliance and clarity of the projected pictures. Other details of the basic method will be found in Engen, Levy, and Schlosberg (1958).

RESULTS

Table 1 presents the distribution of judgments from 390 *S*s together with the mean and average deviation for each of the 48 pictures. There were 234 men and 156 women in the group.

The mean and *AD* values refer to a circular scale, divided into six equal steps numbered *seriatim* as listed under Method. Thus, Picture 1 had a mean of 1.00, which places it exactly in the center of the Love, Mirth, Happiness step, and its very small *AD* shows that almost nobody assigned it to any other category.

TABLE 1

THE DISTRIBUTION OF 48 PICTURES OF FACIAL EXPRESSIONS IN THE SIX CATEGORIES OF THE WOODWORTH SCALE

Picture No.	Categories						Mean	*AD*
	(1)	(2)	(3)	(4)	(5)	(6)		
	Love Mirth Happiness	Surprise	Suffering	Determination Anger	Disgust	Contempt		
1	279	7	2	1	0	1	1.00	.04
2	1	97	220	28	12	32	2.75	1.03
3	69	11	1	37	102	170	5.75	.75
4	102	244	5	8	4	27	1.63	.61
5	12	14	219	51	63	31	3.59	.94
6	0	83	241	40	8	18	2.79	.60
7	2	1	69	130	74	114	4.60	.98
8	147	16	21	125	16	65	.93	1.35
9	2	2	15	24	162	185	5.37	.65
10	0	14	73	262	8	33	3.93	.50
11	1	4	11	216	43	115	4.66	.88
12	10	8	8	21	176	168	5.45	.65
13	205	42	2	28	47	66	.49	.89
14	359	10	0	5	7	9	.93	.21
15	373	16	0	0	0	1	1.04	.08
16	351	19	9	5	2	3	1.12	.23
17	76	26	5	60	64	159	5.84	.90
18	307	75	0	2	2	4	1.19	.35
19	16	304	60	6	1	3	2.14	.31
20	8	3	177	74	97	31	3.89	.94
21	1	4	65	149	87	84	4.48	.91
22	0	1	3	219	29	138	4.77	.98
24	6	9	16	109	109	141	5.10	.84
26	197	102	25	14	21	31	1.34	.78
28	3	36	204	88	28	31	3.49	1.05
29	0	32	184	144	19	11	3.47	.71
30	136	13	4	34	28	175	.19	.73
32	378	7	2	1	1	1	1.02	.02
33	18	3	84	60	167	58	4.65	.87
35	365	19	1	0	4	1	1.03	.08
36	3	2	23	6	294	62	5.05	.34
37	3	6	325	13	16	27	3.31	.58
38	161	199	13	6	3	8	1.69	.59
39	366	18	1	0	0	5	1.04	.07
41	7	10	48	9	244	72	5.05	.58
42	264	48	43	23	5	7	1.21	.78
43	71	246	23	13	11	26	1.72	.63
45	23	64	189	18	53	43	2.72	1.05
46	20	10	19	157	65	119	4.84	1.01
47	4	3	9	310	14	50	4.28	.55
48	226	144	8	3	3	6	1.35	.54
49	67	6	213	9	70	25	3.24	1.07
50	131	225	22	3	3	6	1.70	.55
51	2	27	179	164	3	15	3.47	.68
52	3	4	229	13	117	24	3.80	1.01
54	4	5	78	33	74	196	5.65	1.04
55	196	171	18	1	1	3	1.53	.57
56	96	6	213	13	56	6	2.86	.96

The scale values were computed by the method described by Schlosberg (1941), which avoids the complications introduced by a circular scale through the expedient of rotating the scale so that the modal category has a value of either 3.00 or 4.00 during calculation, and then adding the appropriate correction factor to bring the mean back to its proper place on the original scale.

At the suggestion of Harris Savin, who was assisting in one of the laboratory sections, a more elegant treatment was attempted. Each judgment is treated as a vector of magnitude 1 and angle θ; averaging is carried out in terms of appropriate values of sines and cosines, and the result is converted back to the conventional circular scale value. When the 48 means were recalculated by this cumbersome method, there were no important shifts; 38 of the new means fell ± 0.1 scale step from the values in Table 1, 9 others differed by ± 0.2, and 1 picture (No. 8) differed by 0.6 step. This particular picture, Strain, had the highest *AD* of the series, and actually had two modes on opposite sides of the circle; one cluster of judgments fell in Step 1 (Love, Mirth, Happiness) and the other in Step 4 (Anger, Determination). For such a picture any method of calculating a mean is unsatisfactory.

Sex and sample differences.—The norms given above were on a mixed population gathered over two consecutive semesters. It was of some interest to see if men and women responded differently to the Lightfoot series and to examine the stability of norms from one group to another among Brown University students. There were available the judgments of 390 students, divided between two successive semester courses. Of these, 130 were obtained in Semester I and 260 in Semester II. There were 245 men and 145 women. The data on the first 16 pictures are presented in Table 2. For comparison purposes the modal category and the percentage

TABLE 2

DIFFERENCES BETWEEN SEXES (M AND F) AND BETWEEN SUCCESSIVE GROUPS (I AND II)

Picture No.	Sex Differences				Group Differences			
	Modal Category		Percentage of *Ss* in Modal Category		Modal Category		Percentage of *Ss* in Modal Category	
	M(245)	F(145)	M	F	I(130)	II(260)	I	II
1	1	1	96	98	1	1	97	98
2	3	3	51	65[b]	3	3	55	58
3	6	6	41	48	6	6	42	45
4	2	2	66	58	2	2	62	65
5	3	3	53	62	3	3	59	52
6	3	3	64	58	3	3	66	56
7	4	4	32	39	4	4	34	35
8	1	1	37	39	1	1	36	43
9	6	5	37	47[a]	6	6	45	53
10	4	4	65	71	4	4	66	71
11	4	4	54	58	4	4	56	55
12	5	5	43	50	5	5	46	44
13	1	1	50	57	1	1	71	69
14	1	1	90	95	1	1	93	89
15	1	1	95	97	1	1	95	95
16	1	1	88	94	1	1	90	88

[a] For Picture 9, M-F comparison yields a chi square of 3.58, $df = 1$, $P = .10 > .05$.
[b] For Picture 2, Male-Female comparison yields a chi square of 6.36, $df = 1$, $P = < .01$.

of *S*s placing each picture in the category were considered. The left half of Table 2 shows that the only picture in which there was a sex difference in the modal category was Picture 9, which Table 1 lists with a mean of 5.37, very close to the value of 5.5 which separates Category 5 from 6. This picture, posed as Extreme Pain, seems to have been judged as either Disgust or Contempt (Table 1) and the prediction from the dimensional judgments (see below) put it in the middle of the Disgust category. The other picture to show a sex difference was No. 2, which was posed as Anxiety. Both sexes put it in Category 3 (Fear, Suffering), but women seemed to be significantly more able to recognize it. It is of interest to note that while this was the only picture which yielded statistically reliable differences between men and women judges, 14 of the 16 differences showed higher agreement among the women judges.

The right half of Table 2 shows that there was complete agreement on modal categories by two successive crops of students, and that there were no significant differences in the percentages of students selecting the modal categories. Hence we may assume that women show a slight tendency to excel men in their ability to recognize occasional facial expressions in the Lightfoot series, representing a female face. (Somewhat more pronounced differences have since been found on the Frois-Wittmann pictures of a male face. These will be reported in a later paper.)

Effects of short exposures.—The standard method was to present each picture for 10 sec., permitting *S*s to see it while they were recording their judgments. It seemed desirable to determine the effect of forcing *S*s to make quick judgments. Hence 150 new *S*s were give .5-sec. exposures of the pictures and allowed the remaining 9.5 sec. of the interpicture interval to record their judgments. The results will not be presented in detail, for they showed no major or systematic shifts in means, while the *AD*s increased consistently by less than 10%.

Prediction of Woodworth scale values from dimensional ratings.—Schlosberg (1952) reported that surprisingly good predictions of the Woodworth scale values could be made from independent ratings on two dimensions, Pleasantness-Unpleasantness, and Attention-Rejection. He reported correlations of .94, .92, and .96 between predicted and obtained scale values, in two independent experiments with the Frois-Wittmann pictures and one with the Ruckmick series. But a more revealing measure was the average error of prediction, which was of the order of a half-scale step where chance would be 1½ steps. The average error of prediction of the present scale positions based on predictions from the P-U and A-R dimensional ratings given in Engen, Levy, and Schlosberg (1958) is .60 scale steps. The error can be cut to .50 scale steps by shifting the arbitrary origin of the ordinate to .60 instead of 1.00 on the Woodworth scale; the same adjustment cut the Ruckmick average error of prediction to .42 steps. It may be more than coincidence that rotation helps prediction for both series of female pictures, but not for the Frois-Wittmann male series (Schlosberg, 1952).

The dimensional ratings do not seem to predict the Woodworth scale values quite as well in the Lightfoot series as they did in the two earlier series reported by Schlosberg (1952), although they are of the same order of magnitude. However, this should have been ex-

pected, for the Lightfoot series was deliberately planned to include some pictures of low Sleep-Tension value, where the other two dimensions have restricted range. This should yield more pictures that are near the neutral position, and hence harder to judge in the circumferential dimension described by the Woodworth scale.

Summary

Norms are presented for the Woodworth scale positions of the 48 pictures constituting the Lightfoot series of facial expressions. The norms are quite stable, showing little change with sex of judges, method of calcution, duration of exposure, or from successive groups of judges. However, there is some indication that women are better than men in judging a few of the pictures.

The present study confirms the earlier studies on two other series of facial expressions, in which it was found that the scale positions could be predicted fairly well from independent ratings on the dimensions of Pleasantness-Unpleasantness and Attention-Rejection.

References

Engen, T., Levy, N., & Schlosberg, H. A new series of facial expressions. *Amer. Psychologist*, 1957, **12**, 264–266.

Engen, T., Levy, N., & Schlosberg, H. The dimensional analysis of a new series of facial expressions. *J. exp. Psychol.*, 1958, **55**, 454–458.

Morgan, C. T. *Student's workbook to accompany introduction to psychology.* New York: McGraw-Hill, 1956.

Schlosberg, H. A scale for judgment of facial expressions. *J. exp. Psychol.*, 1941, **29**, 497–510.

Schlosberg, H. A description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44**, 229–237.

Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

Triandis, H. C., & Lambert, W. W. A restatement and test of Schlosberg's theory of emotions, with two kinds of subjects from Greece. *J. abnorm. soc. Psychol.*, 1958, **56**, 321–328.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received August 15, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 126–129

TACTUAL FORM DISCRIMINATION: A DEVELOPMENTAL COMPARISON UNDER CONDITIONS OF SPATIAL INTERFERENCE[1]

EUGENE S. GOLLIN

Queens College

There is evidence which indicates that accurate apprehension of tactual stimulus patterns is disrupted when there is damage to the central nervous apparatus. Similar apprehension difficulties occur among young normal children and in geriatric groups with no manifest neurological signs.

Cohn (1951) has reported that when two cutaneous tactile stimuli are simultaneously applied to ipsilateral surfaces only the more rostral stimulus is perceived by a group of adults with damage primarily in the thalamoparietal projection system. The stimulus applied to the face obliterates stimuli applied simultaneously to the elbow, wrist, or hand. Fifty essentially normal children between the ages of 3 and 10 yr. were also tested. Children between the ages of 3 and 5½ yr. show persistent patterns of response that resemble the behavior of adults with irreversibly disturbed brain function. After the age of 6 yr. the phenomenon of rostral dominance is no longer evident.

Fink and Bender (1953) tested 300 normal children between the ages of 3 and 15 yr. In addition to ipsilateral asymmetric stimulation they also simultaneously stimulated the cheek and the dorsum of the contralateral hand. The child was asked to point to the sites stimulated. They report that 80% of the children under the age of 6 yr. fail to localize both stimuli during the initial 10 trials. In older children the number of failures fell sharply.

Green and Bender (1953) compared normal children aged 3–6 yr. with normal adults by the same methods. In normal adults errors occur only during the initial trials and then disappear. In children the errors persist during many trials. Normal adults over 65 yr. tend to resemble young normal children in the incidence and pattern of perceptual errors.

[1] This research was supported by grants from the National Science Foundation.

Utilizing more complex stimuli (forms constructed of raised tacks) and a method of active manual exploration, Ross (1954) found that normal adults were superior to brain-injured adults in tactual form discrimination. A similar advantage in form discrimination was found for normal adolescents when they were compared with retarded adolescents (Hall, 1956).

The present experment is designed to compare normal children with normal adults on tactual form discrimination in the absence and presence of spatially distributed interferents. In the former task only the "figures" are presented for judgment. In the latter task, judgments are made with the interferents spatially distributed about the form. It is hypothesized that the performance of children on the interferent task (Task B) will be inferior to adult performance; the discrepancy between the groups on the noninterferent task (Task A) will not be as great.

METHOD

Subjects.—All 60 adult *S*s were college students who volunteered their services. The children were pupils in the first three grades of a public elementary school.[2] The summary descriptive data of the 30 children tested on Task A and on Task B are: Mean CA = 90.28 mo., *SD* = 12.28; Mean

[2] The writer wishes to express gratitude to Edward Moses, Principal of Public School 201, Flushing, New York, and to his staff for their cooperation in providing subjects and working facilities for the research. Acknowledgment is also made of the excellent work done by Charles S. Feldstone who served as Research Assistant.

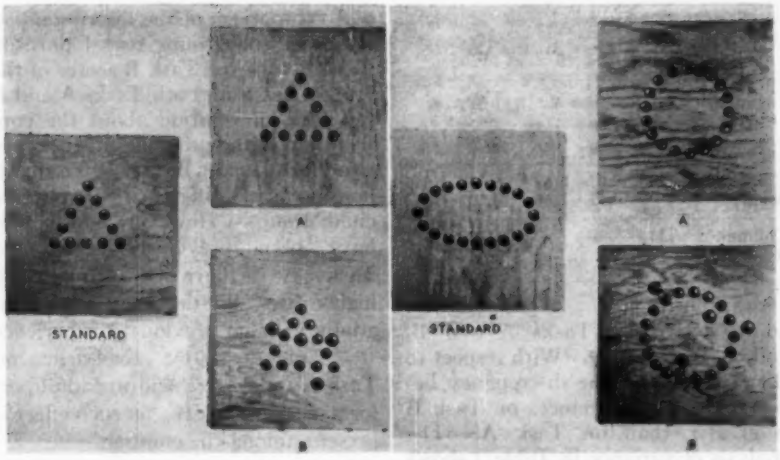

FIG. 1. Examples of materials used. In Task A the standards are compared with the A forms. In Task B the standards are compared with the B forms.

IQ = 121.37, *SD* = 4.75.[3] The 30 children who were tested on the interferent task only had the following descriptive characteristics: Mean CA = 91.50 mo., *SD* = 9.35; Mean IQ = 121.03, *SD* = 11.66.

Materials and procedure.—There were 20 standard forms constructed of metal tacks hammered onto 10 × 10-in. sheets of shellacked plywood. There were 20 comparison forms without interferent tacks and 20 with the interferents present (see Fig. 1). In Task A (no interferent tacks) and in Task B (interferent tacks) the standard was identical with the comparison form in half the trials and different from the comparison form in the balance of the trials. The respective dimensions of the tacks were as follows: small ("figure" tacks), diameter 1.52 cm., height .42 cm.; large ("interferent" tacks), diameter 1.82 cm., height .45 cm.[4]

In both tasks the standard and comparison forms were presented in separate curtained boxes placed side by side. The *S* never saw the tactual forms. The *S* was required to explore the standard form for 15 sec. with his preferred hand, hand flat with fingers together, and was then given up to 15 sec. on the comparison form with the same hand. He was told that the "bumps" on each board make a form or shape, like a circle, for example, and that for each pair of boards he must decide whether the successively explored forms are just the same or different. Before *S* was given Task B he was required to demonstrate that he could distinguish between a small and a large tack. All *S*s who were tested were able to make this discrimination. Prior to Task B, *S*s were told to disregard the larger "bumps" in making their judgments. Thirty children and 30 adults were tested on Task A and approximately 7 days later they were tested on Task B. The same number of children and adults were also tested on Task B only.

Only 15 items were scored. The 5 items not included were found to be too easy in pilot studies: that is, they failed to discriminate either between age groups or experimental conditions. All 20 items were administered in order to maintain the length of the tasks.

RESULTS AND DISCUSSION

Table 1 contains the discrimination score means and *SD* of children and

[3] The IQ scores were obtained from school records and are based on the Pintner-Cunningham Primary Test: Form A.

[4] Appendix A, containing detailed instructions, pictures of the apparatus, and of the forms employed, has been deposited with the American Documentation Institute. Order Document No. 6302, remitting $2.25 for 35-mm. microfilm or $5.00 for 6 × 8-in. photocopies.

TABLE 1

DISCRIMINATION SCORE MEANS AND *SD*s

Groups	*N*	Task A		Task B	
		Mean	*SD*	Mean	*SD*
Children	30	9.07	1.62	6.63	1.56
Adults	30	10.77	1.50	10.70	2.09
Children	30	—	—	6.73	1.72
Adults	30	—	—	9.03	2.70

adults tested on Tasks A and B, and on Task B only. With respect to the A-B groups, the discrepancy between their performances on Task B is greater than on Task A. The analysis of variance in Table 2 shows that Age Levels and Tasks both yield significant *F* values. The interaction term suggests that the spatial interferents contained in Task B are significantly more disruptive for children than for adults. There are statistically significant differences between children and adults on both Task A and Task B. The *t* value for the former comparison is 4.23, $P < .001$; for the latter comparison $t = 8.50$, $P < .001$. The difference between the children's means on Tasks A and B is statistically significant ($t = 5.89$, $P < .001$). There is no significant difference between the A and B performances of the adults.

TABLE 2

ANALYSIS OF VARIANCE OF DISCRIMINATION SCORES

Source	*df*	*MS*	*F*
Age Groups	1	249.41	85.41*
Tasks	1	46.88	16.05*
Age × Tasks	1	42.00	14.38*
Within	116	338.50	
Total	119		

* $P < .001$.

A comparison of the discrimination scores of the groups tested only on Task B with the Task B scores of the groups tested on both Tasks A and B provides information about the contribution of the previous A experience to performance on B. There is a slight, nonsignificant difference between the child groups. However, adults who have been previously tested on Task A achieve a significantly higher discrimination mean score than adults with no previous A experience ($t = 2.68$, $P < .01$). Experience on Task A appears to facilitate adult performance on Task B—no such effect is present among the children.

The results indicate that the children in this investigation were less able than the adults to disregard the spatially distributed interferents. The apprehension of the forms is differentially impaired for the children by the spatially placed large tacks. In the rostral dominance studies reported above the young children appeared to have difficulty in apprehending simultaneously applied tactile stimuli. In the present investigation a somewhat older group has difficulty in disregarding tactile stimuli during an attempt to apprehend a configuration constructed of discriminably smaller tactile stimuli.

SUMMARY

Children and adults were compared on two versions of a tactual discrimination test. In the first version they were required to indicate whether manually explored pairs of forms were the same or different. In the second task they were required to make the same comparison with background interferents scattered among one member of each pair. Adults were more successful than children under both conditions; their success under the second condition was considerably greater. Adults who were tested on the two tasks performed significantly better on the second task than a group of adults who were only tested on the second task. No such differ-

ences were found between comparably tested groups of children.

References

COHN, R. On certain aspects of sensory organization of the human brain: II. A study of rostral dominance in children. *Neurology*, 1951, **1**, 119–122.

FINK, M., & BENDER, M. B. Perception of simultaneous tactile stimuli in normal children. *Neurology*, 1953, **3**, 27–34.

GREEN, M. A., & BENDER, M. B. Cutaneous perception in the aged. *AMA Arch. Neurol. Psychiat.*, 1953, **69**, 577–581.

HALL, A. J. Comparative study of the performance of normal and subnormal adolescents on a tactual perception test. Unpublished master's thesis, Syracuse Univer., 1956.

ROSS, A. O. Tactual perception of form by the brain-injured. *J. abnorm. soc. Psychol.*, 1954, **49**, 566–572.

(Received August 15, 1959)

Journal of Experimental Psychology
1960, Vol. 60, No. 2, 130–136

EFFECT OF SELF-SATIATION ON PERCEIVED SIZE OF A VISUAL FIGURE[1]

CARL P. DUNCAN

Northwestern University

In Köhler and Wallach's (1944) study of visual figural aftereffects (FAE) a few cases are reported in which the contour of the test (T) figure was identical to and coincident with the contour of the previously fixated inspection (I) figure. The usual finding in these cases was that, among other FAEs, the test figure seemed to appear smaller in size than a comparison figure of the same objective size. Hebb (1949), McEwen (1958), and Sagara and Oyama (1957) claim that this finding is a serious, perhaps the most serious, objection to Köhler and Wallach's neural satiation theory. The theory deals primarily with contour displacement, within which the size effect described above is merely one variety. According to the authors just cited, the theory should predict no displacement, no size change, when inspection and test figures have identical contours and these contours fall in exactly the same place in the visual field.

There now seems no question that the T figure appears smaller in the case of identical and coincident T and I contours. Sagara and Oyama (1957, Fig. 2) summarize six studies, and there are at least two others (Mori, 1956; Oyama, 1959), all of which found that the T figure appeared reduced in size in this case.

All of the eight Japanese studies just cited used outline circles both as I and T figures, and all measured FAE, i.e., change in size of the test circle. Even though the circles may have been identical, and the time interval between them only a fraction of a second (though it was sometimes longer), one cannot determine from these studies whether there were any changes in the I circle itself during the inspection period (Köhler and Wallach's [1944] "self-satiation"). Since self-satiation effects are not necessarily the same as FAEs (Köhler & Wallach, 1944, p. 348f.), it seems worthwhile to determine whether an outline I circle alone shows any apparent change in size after it has been fixated for some time. This is the purpose of the present study.

An experiment similar in basic procedure to the present study has been reported as part of a larger investigation of self-satiation by Ikeda and Obonai (1953). The same five *S*s served in 12 conditions in their experiment, fixation to the left or the right of an outline circle (I figure) for 1, 15, 30, 60, 120, or 240 sec. During fixation, only the I circle was visible; a comparison circle on the other side of the fixation point was covered by a white card. At the end of inspection, the card was removed and *S* judged the size of the I circle against the comparison circle. The findings were that the I circle appeared reduced in size. The shrinkage increased to a maximum at about 60-sec. inspection time, after which there was no further change. Since

[1] This investigation was supported by Research Grant B-792 from the National Institute of Neurological Diseases and Blindness of the National Institutes of Health, United States Public Health Service. Thanks are due Bruce O. Bergum and Keith N. Clayton for assistance in the study.

the shrinkage was quite small, and since the findings were based on relatively few *Ss* who served in all conditions, it seemed desirable to supplement Ikeda and Obonai's experiment.

METHOD

Apparatus.—A white posterboard screen, 40 in. wide, 22 in. high, was mounted vertically on a wooden frame attached to the front edge (as viewed by *S*) of a table. An aperture, 9 in. wide and 7 in. high, was cut in the screen and centered in the screen in the horizontal dimension. The top edge of the aperture was 5 in. from the top of the screen; its center was 43 in. from the floor. The aperture was filled with clear Plexiglas on the front surface of which was painted a small black fixation cross centered in the aperture. A white posterboard shutter could be raised or lowered by *E*, sitting behind the screen, to uncover or cover the aperture.

The *S* fixated the fixation mark from a sitting position at another table on the edge of which was mounted a Bausch and Lomb adjustable head-and-chin rest. When *S*'s head was fixed in the rest the distance from the cornea to the fixation mark was normally 12 ft. 6 in., but *S* was allowed to tip the rest forward or backward by as much as 1 in. to obtain the most comfortable position for him.

Stimuli.—The I stimulus was an outline circle 2 in. in outside diameter. There were nine comparison circles varying in diameter, by steps of 1/16 in., from 1 12/16 in. to 2 4/16 in. The middle comparison stimulus was the same diameter as the I circle. The contour was 3/32 in. wide in all circles.

All stimuli were drawn in India ink, photographed, reproduced as glossy positive prints (black circle on white background), and drymounted on posterboard cards backed by Masonite to prevent warping. In addition to the cards bearing the I and comparison stimuli, a blank card (glossy positive paper mounted the same as the stimuli) was also available. All cards were slightly taller than the vertical size (7 in.) of the aperture to permit them to slide easily, but be held firmly, in tracks mounted on the rear of the posterboard screen above and below the Plexiglas window. Since the cards were presented behind, and pressed firmly to, the Plexiglas window, the fixation cross and the stimuli differed by a distance of ¼ in., the thickness of the Plexiglas.

The card bearing the I circle was mounted permanently behind the window and was never moved during the course of the experiment. As viewed by *S*, the I circle was centered vertically in the window to the right of the fixation cross such that the center of the circle was in the same horizontal plane as the cross. At its closest point, the left outside edge of the contour of the circle was ½ in. to the right of the cross. The left vertical edge of the card bearing the I circle passed through the cross. The blank card, or the cards bearing the comparison circles, could be slipped into position to the left of the cross such that their right vertical edges also passed through the cross. Edges of cards were trimmed smooth and straight by a power papercutter, so the line formed by adjoining edges of the inspection card and any other card, and passing through the fixation cross, was nearly invisible from *S*'s position 12½ ft. away. When a card bearing a comparison circle was in place, the right outside edge of the circle's contour was, at its closest point, ½ in. to the left of the fixation cross; this distance remained constant for all sizes of comparison circles. The total pattern formed by I circle, fixation cross, and comparison circle, subtended about 2° of visual angle. Under the room illumination used throughout the experiment the luminance of the white background of a circle (glossy positive paper), measured through the Plexiglas window by MacBeth illuminometer, was about 16 mL.

Conditions.—There were three conditions: control, continuous inspection, and distributed inspection. In the control condition, *Ss* judged the size of the I circle against the comparison circles without a prior period of fixating the I circle. In the continuous inspection group, *Ss* first fixated the fixation mark (I circle on the right, blank card on the left) continuously for 2 min., at the end of which time they made judgments in the same way as the control group. Since *Ss* find it difficult to maintain fixation continuously for periods as long as 2 min., and since some of them report qualitative changes in the stimulus during such periods (Marks, 1949), a distributed inspection group was run in the same way as the continuous group except that the 2 min. of fixation was presented in cycles of 15 sec. fixation and 10 sec. rest. At the end of the last (eighth) 15-sec. period these *Ss* made judgments in the same way as the other two groups.

Subjects.—The *Ss* were 540 students from elementary psychology courses, 180 in each of the three conditions. When students were asked to serve as *Ss* for the experiment it was specified that "only those with fairly good vision or good vision corrected with glasses

should serve." During the experiment all *Ss* were questioned on whether they could easily see the fixation cross from their viewing position; the records of two *Ss* who reported difficulty seeing the cross were discarded. The records of seven *Ss* who failed to follow instructions or for whom *E* made procedural errors were also discarded. All rejected *Ss* were replaced.

Procedure.—As soon as *S* was seated comfortably with his head in the headrest, *E* explained that when the shutter was raised *S* would see a small black cross. It was strongly emphasized to *S* that whenever the cross was in view at any time during the experiment *S* must look directly at the cross and never move his eyes from it (binocular viewing was used throughout). The *S* was told he could, and should, blink but always to fixate the cross whenever it was present. Then *S* was told that to the right of the cross there would be a circle which would remain in position throughout the experiment, that *S*'s task would be to judge the size of this circle against other circles which would be presented to the left of the cross, and that these judgments would have to be made while fixating the cross. The *S* was told that each time a new circle appeared on the left he would have a few seconds during which he must say whether the circle on the left was larger or smaller than the fixed circle on the right, no equal judgments allowed.

The *Ss* in the continuous inspection group were told that before they made these judgments there would be a period during which they would simply fixate the cross and that circles would be presented for judgment at the end of this period. The *Ss* in the distributed inspection group were given this instruction and were also told that during the time they were fixating the cross the shutter would be dropped to cover the window occasionally. These *Ss* were told that while the shutter was down they were not to look at anything but the white surface of the shutter.

As soon as *S* understood what he was to do, the shutter was raised and *E* simultaneously said, "Fixate the cross," to all *Ss*. For *Ss* in the control group a comparison circle was already in place when the shutter was raised, and these *Ss* were immediately told to say, without looking away from the cross, whether the circle on their left was larger or smaller than the one on their right. Several times during the 2-min. inspection period for the continuous group these *Ss* were told to blink but keep fixating the cross. The *Ss* in the distributed inspection group were told to fixate the cross each time the shutter was raised at the end of a 10-sec. rest period. During the last 15 sec. of inspection, *Ss* in both the continuous and distributed inspection groups were warned that soon *E* would start presenting other circles to the left of the cross which were to be judged as larger or smaller than the circle on the right.

Each of the nine comparison stimuli was presented once to each *S*. Each comparison stimulus appeared for 4 sec. In order to minimize satiation of the comparison stimuli, a 10-sec. rest was given after presentation of each stimulus. For the control group, the shutter covered the window during these 10-sec. periods to prevent accumulation of satiation to the inspection circle. For the continuous and distributed inspection groups, the blank card was put back into position after presentation of each comparison stimulus and the shutter was raised during the 10-sec. rest periods, with instructions to fixate, to prevent dissipation of satiation to the inspection circle. Thus, total fixation time for both inspection groups was 2–4 min. by the end of the experiment.

The nine comparison stimuli were arranged in nine completely different orders of presentation so that each stimulus appeared once as first stimulus presented, once as second stimulus presented, etc. Within each of the three groups, 20 *Ss* went through each of the nine orders.

Results

The percentage of *Ss* (out of 180) in each group that judged each of the nine comparison stimuli larger than the I stimulus was determined and these percentages were used to obtain, by the method of least squares (Woodworth & Schlosberg, 1954, p. 206), a mean size of comparison circle judged equal to the I circle. These means (and *SD*s) are shown in Table 1. It may be seen that both the continuous and distributed inspection groups have means smaller than the mean of the control group, suggesting some reduction in apparent size of the I circle.

Following the suggestions of Woodworth and Schlosberg (1954, p. 208), the *N* used for *t* tests between the

TABLE 1

MEAN SIZE OF CIRCLE IN INCHES JUDGED EQUAL TO THE 2-IN. INSPECTION CIRCLE (One set of means is based on all comparison stimuli, the other on only the first-presented comparison stimulus)

Group	All Stimuli		First Stimulus	
	Mean	*SD*	Mean	*SD*
Continuous	1.9902	.0699	1.9837	.0729
Distributed	2.0061	.0642	1.9854	.0708
Control	2.0244	.0807	2.0078	.0708

means of Table 1 was simply the number of *S*s in each group, 180. The control mean was significantly larger than either the mean of the continuous inspection group ($t = 4.28$, $P < .001$) or the mean of the distributed inspection group ($t = 2.37$, $P < .02$). The continuous inspection mean was smaller than the distributed inspection mean at a low level of significance ($t = 2.24$, $P < .05$).

Table 1 also shows means computed from judgments of only the first comparison stimulus presented to *S* (each of the nine comparison stimuli appeared once as first stimulus for 20 *S*s in each group). As can be seen, these means, which are less likely to be influenced by any changes in satiation that might occur during the testing period, yield the same rank order as the means based on all stimuli. Again using $N = 180$ for each group, the control mean was significantly larger than either the continuous group mean ($t = 3.17$, $P < .01$) or the distributed mean ($t = 2.99$, $P < .01$). The continuous and distributed means did not differ ($t = .22$). Thus, either with all stimuli or with first stimulus alone, the size of circle judged equivalent to the inspection circle was smaller for either inspection group than for the control group. The distributed group, in which some dissipation of satiation could occur during the 10-sec. rest intervals interspersed in the fixation period, tended to show less reduction in size of the circle than the continuous group.

In a previous study (Duncan, 1958) it was found that a different measure yielded results highly correlated with those based on means computed by least squares. This measure was the number of larger judgments made by each *S*. It has the advantage that one obtains directly a raw score for each *S*, thereby permitting use of analysis of variance. In the present experiment each *S* judged nine stimuli once each, thus *S*'s score could range from 0 to 9. Table 2 shows the mean number of larger judgments for each group. Since a higher mean value indicates that more of the comparison stimuli were judged larger than the I circle, the pattern of these means yields the inverse rank order of the three groups as obtains in the means of Table 1, i.e., Table 2 also suggests that the continuous inspection group appears to have perceived the inspection circle as smaller than did the distributed inspection group, which in turn perceived the circle as smaller than did the control group.

In the analysis of variance on the means of Table 2, opportunity was taken to test for any differences among the nine orders in which the comparison stimuli were presented. Thus, a two-way analysis was done in which major group was one variable, stimulus order the other. The analysis revealed highly significant differences among Groups ($F = 18.90$, $P < .001$, 2/513 *df*). Stimulus Orders ($F = 1.76$, 8/513 *df*) and the interaction ($F = 1.01$, 16/513 *df*) were not significant. By *t* test, the control group mean was significantly different from both the continuous ($t = 5.78$) and the distributed ($t = 4.44$) inspection groups. The continuous and distributed groups did not differ ($t = 1.51$). Thus, this measure, number of larger judg-

TABLE 2

MEAN NUMBER OF LARGER JUDGMENTS

Group	Mean	*SD*
Continuous	4.661	.895
Distributed	4.522	.840
Control	4.117	.885

ments, also suggests that the inspection circle appeared smaller to both inspection groups than it did to the control group.

Discussion

The results confirm the findings of Ikeda and Obonai (1953); after prolonged fixation at the same place in the visual field, an outline circle appeared slightly reduced in size. The shrinkage was quite small. With a circle subtending about 34′ of visual angle, Ikeda and Obonai found a maximum reduction of about .8′ (.67 mm. shrinkage in a 30-mm.-diameter circle) after 1–4 min. inspection. In the present study, comparison of the continuous inspection group and control group means based on all stimuli (Table 1) indicates a shrinkage, after 2–4 min. inspection, of about .8′ in a circle subtending about 46′ (.0342 in. shrinkage in a 2-in.-diameter circle). Data from the present study may include some space error, since the inspection circle was always to the right of the fixation point, but in Ikeda and Obanai's experiment space error was controlled.

It may be of interest to compare with these values the amounts of shrinkage obtained when a T circle is coincident with an I circle, i.e., in the figural aftereffect case that most closely corresponds to the case of self-satiation of the I circle. Oyama (1956) gives the parameters needed to determine the shrinkage found by the six studies summarized in Sagara and Oyama's (1957) Fig. 2 (the necessary data are not available for any other studies). Table 3 gives the size of the T circle (which is also the size of the I circle) in visual angle, the time of fixation of the I circle, and the shrinkage of the T circle (computed from the percentage shrinkage reported by Sagara and Oyama). The labeling of the studies is the same as that used in Sagara and Oyama's Fig. 2. It can be seen that five of the six studies found shrinkage ranging from slightly less than 1′ to slightly more than 2′. Kogiso found shrinkage of more than 6′, but his circle was considerably larger than the circles used by the other five investigators. Ikeda and Obonai also used a fairly large circle and found the smallest amount of shrinkage, but their inspection time was only .5 sec.

TABLE 3

Apparent Shrinkage in Visual Angle of the T Circle Coincident with the I Circle, Original Size of the T Circle, and Inspection Time for the Six Studies Reported by Sagara and Oyama (1957, Fig. 2)

Study	Inspection Time (Sec.)	Test Circle Size	Shrinkage
Oyama 1	15	22.9′	1′
Oyama 2	15	45.8′	2.4′
Oyama 3	15	45.8′	1.9′
Ikeda	120	34.4′	1.8′
Ikeda & Obonai	.5	1° 43.1′	.8′
Kogiso	15	2° 41.7′	6.5′

The FAE studies in Table 3 that afford the best comparison to the self-satiation experiments (i.e., the present study, and Ikeda & Obonai, 1953) are the three identified in the table as Oyama 2, Oyama 3, and Ikeda. The two Oyama experiments used circles essentially identical in size to that used in the present study (nearly 46′). Ikeda's circles (about 34′) were essentially identical to that used by Ikeda and Obonai (1953). Inspection times were probably long enough in all five studies to achieve nearly maximal amounts of shrinkage. Table 3 shows that the three FAE studies all found roughly 2′ shrinkage of the T circle. As reported earlier, the two self-satiation experiments found about .8′ shrinkage of the I circle. It does not seem likely that the difference is reliable, but at present this cannot be determined. (In percentage shrinkage, Ikeda and Obonai [1953] found 2.2% shrinkage due to self-satiation; the value for the present study was 1.7%. As may be seen in Sagara and Oyama's [1957] Fig. 2, all of the figural aftereffect studies except Ikeda and Obonai [who used only .5 sec. inspection] found shrinkage between 4% and 5.5% when test and inspection circles coincided.)

If the findings of the present study

and those of Ikeda and Obonai (1953) on self-satiation are reliable, then neural satiation theory is confronted with a size change or contour displacement that seems at least as difficult, perhaps more difficult, to account for than is the FAE shrinkage of coincident test figures. Hebb (1949, p. 56) believes the theory should predict that the perceived size of a square or circle should increase, not decrease, as satiation increases because, as Köhler and Wallach (1944) repeatedly insist, satiation is greater inside than outside the contour of a closed outline figure. Contours should recede from more highly to less highly satiated areas. The more specific version of neural satiation theory presented by Köhler, Held, and O'Connell (1952) seems no better able to account for size reduction due to self-satiation.

Köhler and Wallach (1944, p. 351) seem to imply that the outline circle may be a special case, as regards its own self-satiation. Although, they say, satiation is greater inside than outside the contour of a circle, the satiated area is greater on the outside. But shrinkage due to self-satiation is not limited to outline circles. Ikeda and Obonai (1953) presented two vertical parallel lines on one side of a fixation point. After some fixation, presentation of a comparison figure on the other side of the fixation mark revealed that the fixated lines appeared closer together, and the lines themselves appeared shorter. Nozawa (1955) presented a horizontal line 100 mm. long on one side of the fixation point. After 40 sec. inspection the line appeared shortened by an average (for 9 *Ss*) of 5.67 mm. in comparison with lines of varying length presented on the opposite side of the fixation mark. (Constant errors in judgment of length were controlled.) Increasing the inspection time, or presenting the 100-mm. line intermittently (e.g., at 1 or 2 cycles/sec.) during fixation, increased the apparent shortening of the line to as much as 11 mm. Although Köhler and Wallach (1944, p. 351) say only that self-satiation cannot "displace" or "distort" a straight line, it would seem they should also predict that apparent reduction in its length should not occur.

Despite the fact that a number of studies, both of self-satiation and of FAE, seem to present difficulties for the contour displacement aspect of satiation theory, these difficulties may not be insuperable. There are a number of factors to consider. First, a T figure presented in the same place as the just-removed I figure also shows color and depth changes; the test figure tends to appear grayer and farther away than comparison stimuli presented in relatively unsatiated areas. Since color and depth changes tend to be maximal when the T figure is coincident with the I figure, whereas contour displacement is maximal when the contours of the two figures are separated by a short distance (the distance paradox), Köhler and Wallach did not appeal to color and depth changes to explain displacement. But it is possible that either an I figure alone, in the case of self-satiation, or a T figure, may appear smaller just because it appears grayer (less contrast) and farther away than the comparison stimuli against which it is being judged. The depth effect, which might have the same consequences for judgments of apparent size as an actual reduction in size of the retinal image, would be particularly important to examine in this connection.

Second, when a figure is presented on one side of a fixation mark, some parts of the figure are closer to the fixation mark than are other parts, and the former parts are also more foveally located than are the latter. Since the fixation mark is a visual figure which should also produce satiation, it may have some effect on size changes, particularly on the shortening of Nozawa's (1955) straight horizontal line. Foveal-to-peripheral differences may also produce inhomogeneities of satiation which could influence size judgments (though Köhler and Wallach (1944) and Hebb (1959) disagree on the nature of these inhomogeneities). Also, Köhler and Wallach report that satiation is usually

stronger in the lower half than in the upper half of the visual field. Inhomogeneities of satiation, from any source, may influence size judgments directly, or indirectly by producing changes in a figure's apparent color, distance, or shape.

At present, there are a number of contour displacements which do not readily follow from neural satiation theory. These include some self-satiation effects, and FAEs both in the case of superimposed T and I contours, and in certain other cases (e.g., Köhler & Wallach, 1944, Fig. 19; Smith, 1954). But in view of the theory's many successful predictions, the difficulties that these cases present should not be overstated.

Summary

To test for apparent changes in size of an outline inspection circle following fixation of the circle for a period of time, three groups of 180 *Ss* each judged the size of the circle against a series of comparison circles of varying size. Prior to judging, one group of *Ss* continuously fixated a point immediately to the left of the inspection circle for 2 min. A second group fixated for 2 min. in cycles of 15-sec. fixation and 10-sec. rest. The third, control, group judged without a prior period of fixation.

The results were that the inspection circle was judged significantly smaller in size by both fixation groups than by the control group. The apparent reduction in size was slightly greater for the group that had fixated continuously than for the group that had rests interspersed in the fixation period. The results were discussed in relation to neural satiation theory.

References

Duncan, C. P. Figural displacement with quasi-circular stimuli. *Percept. mot. Skills*, 1958, **8**, 295–305.

Hebb, D. O. *The organization of behavior*. New York: Wiley, 1949.

Ikeda, H., & Obonai, T. The quantitative analysis of figural after-effects: III. On "self-satiation." *Jap. J. Psychol.*, 1953, **24**, 179–192.

Ikeda, H., & Obonai, T. The studies in figural after-effects: IV. The contrast-confluence illusion of concentric circles and the figural after-effect. *Jap. psychol. Res.*, 1955, **2**, 17–23.

Köhler, W., Held, R., & O'Connell, D. N. An investigation of cortical currents. *Proc. Amer. Phil. Soc.*, 1952, **96**, 290–330.

Köhler, W., & Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. Phil. Soc.*, 1944, **88**, 269–357.

Marks, M. R. Some phenomena attendant on long fixation. *Amer. J. Psychol.*, 1949, **62**, 392–398.

McEwen, P. Figural after-effects. *Brit. J. Psychol., Mongr. Suppl.*, 1958, Whole No. 31.

Mori, T. Figural after-effects and stimulus-organization. *Jap. psychol. Res.*, 1956, **3**, 8–14.

Nozawa, S. On the after-effect by intermittent presentation of inspection figure. *Jap. psychol. Res.*, 1955, **2**, 9–16.

Oyama, T. Temporal and spatial factors in figural after-effects. *Jap. psychol. Res.*, 1956, **3**, 25–36.

Oyama, T. A new psychophysical method: Method of transposition or equal-appearing relations. *Psychol. Bull.*, 1959, **56**, 74–79.

Sagara, M., & Oyama, T. Experimental studies on figural aftereffects in Japan. *Psychol. Bull.*, 1957, **54**, 327–338.

Smith, K. "Attraction" in figural after-effects. *Amer. J. Psychol.*, 1954, **67**, 174–176.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

(Received August 20, 1959)

PSYCHOLOGICAL MONOGRAPHS

Coming in September . . .

No. 497. **A Re-evaluation of the Meaningfulness of Possible CVC Trigrams**

By E. JAMES ARCHER Price $1.50

In this study association values for 2480 trigrams (nonsense syllables) were determined, and two measures of reliability were obtained.

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION, 1333 16th St. N.W. Washington 6, D. C.